University of Alberta

The molecular genetics of the *raspberry* locus in *Drosophila melanogaster*

by

Song Hu ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of
the requirements for the degree of Doctor of Philosophy

in

Molecular Biology and Genetics

Department of Biological Sciences

Edmonton, Alberta

Fall 1998

National Library of Canada

Acquisitions and Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada

Acquisitions et services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-34780-X

Canada

University of Alberta

Library Release Form

Name of Author: Song Hu

Title of Thesis: The molecular genetics of the *raspberry* locus in *Drosophila melanogaster*

Degree: Doctor of Philosophy

Year this Degree Granted: 1998

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Song Hu

604C Michener Park
Edmonton, Alberta
Canada T6H 5A1

Aug. 31, 1998

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "The molecular genetics of the *raspberry* locus in *Drosophila melanogaster*" by Song Hu in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Molecular Biology and Genetics.

Dr. D. Nash
(Supervisor)

Dr. J.B. Bell

Dr. K.L. Roy

Dr. M.A. Walter

Denise Clark
Dr. D.V. Clark

Supervisor writes the date that the thesis is approved by committee here July 31 1998

Abstract

The *raspberry (ras)* locus is in region 9E1-E4 on the X chromosome of *Drosophila melanogaster*. Three groups of mutations exist at the locus: spontaneous *ras* eye-colour mutants, defective in pteridine biosynthesis; purine auxotrophs (*gua1*^{ts}, *pur1*$^{1\ \&\ 2}$); and recessive lethal mutants (*ras-l*). They constitute the "*ras* complex". The *ras-l* mutants form a single lethal complementation group. A majority of the *ras-l* mutants fail to complement both the eye-colour and the auxotrophic mutants, implying all three groups are functionally related. However, the *ras* mutants complement *gua1*^{ts} and the *pur1* mutants; in addition, *gua1*^{ts} complements *pur1*^{2}.

Wild-type *ras* DNA was cloned, utilizing mutant DNA previously isolated after P-element tagging the gene (Leonard, 1986). Its DNA sequence encodes inosinate dehydrogenase (IMPDH), the rate-limiting enzyme in GMP biosynthesis. The probable function of the gene is thus in GMP biosynthesis. All mutant phenotypes in the *ras* complex can be rationalized as results of altered IMPDH activity. Several cDNAs were isolated; the longest, which is close to the size (2.4kb) of the main transcriptional product of the region, was sequenced.

The gene contains four introns. There is no TATA-box. Instead, there is an initiator-like element and an extremely high GC content sequence upstream. The combination of initiator and GC rich sequences is typical of house-keeping regulation. The developmental profiles of transcript and IMPDH activity were studied. The highest transcript level occurs in the 0-2 hour embryonic period. These transcripts are probably maternal products; extraordinarily, their absence from *ras*$^{1\ \&\ 2}$ early embryos suggests they are dispensable.

Two *cis*-acting regulatory sequences seem to have been identified in studies of *ras* mutants. One, probably a splicing enhancer, is in the second intron. Insertions in this element generate the classical *ras* eye-colour mutants, *ras*$^{1, 2 \& 4}$. The other is located upstream of the transcriptional initiation site. A deletion in the region generates the male specific mutant *ras*^{3}, suggesting involvement in dosage compensation. If house-keeping and eye-specific functions of the IMPDH gene are independently regulated, tissue-specific regulatory elements seem to be extremely limited; in contrast, the regulatory apparatus concerned principally with house-keeping function is probably more extensive.

Acknowledgements

I am deeply indebted to my supervisor Dr. D. Nash for his help in guiding my research with his helpful advice, discussion, criticism and financial assistance during both the experimental work and thesis preparation.

I also wish to acknowledge excellent advice and technical help from Dr. S. Tiong, Ms. E. Woloshyn and Mr. E. Chomey. I wish to thank Dr. K. Roy particularly for his extensive advice concerning experiments involving enzymatic activity analysis; as well as his role, along with Dr. J. Bell on my supervisory committee. Thanks also to Dr. J. Lock, Mr. W. Clark and G. Ritzel and to the many other people who helped me directly or indirectly.

Finally, I am grateful to my family, especially my wife Jin-Xia and my son Jing, for their encouragement, support and patience throughout the long process leading to the production of this thesis.

Significant segments of *Section 3.2, Section 3.3.1* are derived from "The *raspberry* locus of *Drosophila melanogaster* includes an inosine monophosphate dehydrogenase like coding sequence" by Nash, D., S. Hu, N. J. Leonard, S. Y. Tiong, and D. Fillips in *Genome* 37 (2): 333-44, 1994.

Table of Contents

1 Introduction 1
1.1 Gene regulation and purine metabolism 1
1.2 Gene regulation 2
1.2.1 The natural history of gene expression 2
1.2.2 Transcriptional regulation in bacteria 3
1.2.2.1 Regulation results from DNA/protein and protein/protein interactions 3
1.2.2.2 Transcriptional control regions 3
1.2.2.3 Regulation of transcriptional initiation 5
1.2.3 Transcriptional regulation in eukaryotes 7
1.2.3.1 Prokaryotic and eukaryotic transcriptional systems 7
1.2.3.2 The basics of eukaryotic transcriptional initiation 8
1.2.3.3 Transcriptional activation in eukaryotes 10
1.2.4 Regulation of house-keeping and tissue-specific genes 12
1.2.4.1 Structure and regulation of house-keeping genes 12
1.2.4.2 Tissue-specific gene expression 13
1.3 Introns 15
1.3.1 Intron structure 15
1.3.2 Spliceosome assembly 16
1.3.3 Splice site selection 17
1.3.4 Alternative splicing 17
1.3.5 Introns and transcriptional initiation 18
1.3.6 *Sex-lethal* in Drosophila 19
1.4 Purine metabolism as a model system for studying gene regulation 22
1.4.1 An introduction to purine metabolism 22
1.4.1.1 Purines 22
1.4.1.2 *de novo* purine nucleotide biosynthesis 23
1.4.1.3 Purine salvage pathways 23
1.4.1.4 Purine interconversions 23
1.4.1.5 Purine catabolism 27
1.4.1.6 Rate limiting steps in purine biosynthesis 27
1.4.2 Purine biosynthesis in bacteria 27
1.4.2.1 *Escherichia coli* 27
1.4.2.2 *Bacillus subtilis* 29
1.4.3 Purine biosynthesis in eukaryotes 30
1.4.3.1 *Saccharomyces cerevisiae* 30
1.4.3.2 *Drosophila melanogaster* 31
1.4.3.3 *Homo sapiens* 33
1.5 Drosophila purine genetics 34
1.5.1 Purine metabolism in Drosophila 34
1.5.1.1 Purine biosynthesis 34
1.5.1.2 Pteridine metabolism in Drosophila 35
1.5.2 Auxotrophs in purine biosynthesis 37
1.5.2.1 Auxotrophs in *Drosophila melanogaster* 37
1.5.2.2 Adenosine requiring mutants 37
1.5.2.3 Guanosine requiring mutants 38
1.5.2.4 Purine requiring mutants 38
1.5.3 The *raspberry* complex 38
1.5.3.1 *raspberry* eye-colour mutants 38
1.5.3.2 The relationship of *gua1*, *pur1* and *ras* 39
1.5.3.3 Molecular analysis of the *ras* locus 39
2 Materials and Methods 41
2.1 Stocks 41
2.1.1 *Drosophila melanogaster* stocks 41
2.1.2 Bacteria, bacterial plasmid vectors and bacteriophage λ vectors 41
2.1.2.1 Bacteria: 41
2.1.2.2 Bacterial plasmid vector pUC19 41

2.1.2.3 Bacterial phagemid vector pBluescript ... 44
2.1.2.4 Bacterial phage vector EMBL-4 ... 44
2.2 Media and solutions ... 44
2.2.1 Drosophila medium ... 44
2.2.2 Bacterial media ... 46
2.2.2.1 Luria-Bertani (LB) Broth: ... 46
2.2.2.2 Tryptone (T) Medium: ... 46
2.2.2.3 2YT MEDIUM: ... 46
2.2.3 Stock reagent solutions ... 46
2.3 Tissue collection ... 48
2.3.1 Embryos and earlier larval samples ... 48
2.3.2 Late larvae and pupae ... 48
2.3.3 Adults ... 48
2.4 DNA manipulation ... 50
2.4.1 DNA extraction ... 50
2.4.1.1 Drosophila genomic DNA ... 50
2.4.1.2 Large scale plasmid DNA purification ... 50
2.4.1.3 Small scale plasmid and phage DNA extraction ... 51
2.4.1.4 Single strand phagemid DNA (ssDNA) extraction ... 51
2.4.2 Restriction endonuclease digestion ... 51
2.4.3 DNA agarose gel electrophoresis ... 52
2.4.4 Gel Purification ... 52
2.4.5 Preparation of radiolabeled DNA probes ... 52
2.4.6 Southern transfers and hybridization ... 52
2.4.7 Making and screening an EMBL-4 genomic DNA library ... 53
2.4.7.1 Preparation of vector arms ... 53
2.4.7.2 Preparation of genomic inserts ... 53
2.4.7.3 Ligation of the arms and inserts ... 56
2.4.7.4 Library size ... 56
2.4.7.5 Library amplification ... 57
2.4.7.6 Screening the library ... 57
2.4.8 DNA sequencing and computer analysis ... 57
2.4.9 Polymerase Chain Reaction ... 60
2.4.10 Oligo primers ... 60
2.5 RNA manipulation ... 61
2.5.1 RNA extraction and poly-A$^+$ RNA purification ... 61
2.5.2 RNA agarose gel electrophoresis, northern transfer ... 62
2.5.3 Ribo probes ... 62
2.5.4 Hybridization and washing ... 62
2.6 Analysis of inosine monophosphate dehydrogenase activity ... 63
2.6.1 Protein extraction ... 63
2.6.2 Protein quantification ... 63
2.6.3 The assay reaction ... 63
2.6.4 Chromatography ... 64
2.6.5 Phosphorimaging and computer analysis ... 64
3 Results ... 66
3.1 Cloning of the *ras* gene ... 66
3.1.1 The Am$^+$ EMBL-4 library ... 66
3.1.2 Screening the library ... 66
3.1.3 Analysis of the wild type *ras* DNA ... 68
3.1.4 Isolation of *ras* cDNA clones ... 74
3.2 Sequence analysis of the *ras* gene ... 77
3.2.1 Exon/intron structure ... 77
3.2.2 Transcription and translation start and termination sites ... 77
3.2.3 A Trinucleotide repeat sequence ... 84
3.2.4 Inosine monophosphate dehydrogenase (IMPDH) ... 84
3.3 The molecular lesions in *ras* mutant DNA ... 88
3.3.1 Southern analysis ... 88

- 3.3.2 Cloning of the insertion point containing fragments in *ras*^{1}, *ras*^{2} and *ras*^{4} ... 91
  - 3.3.2.1 *ras*^{1} ... 91
  - 3.3.2.2 *ras*^{2} ... 91
  - 3.3.2.3 *ras*^{4} ... 91
- 3.3.3 Analysis of the insertion points ... 92
  - 3.3.3.1 *ras*^{1} ... 92
  - 3.3.3.2 *ras*^{2} ... 92
  - 3.3.3.3 *ras*^{4} ... 100
- 3.3.4 Analysis of *ras*^{3} mutation ... 100
- 3.3.5 *ras* revertants ... 103
- 3.3.6 Expression pattern of the *ras* gene ... 103
  - 3.3.6.1 Enzyme activity in the late stages of wild-type development ... 103
  - 3.3.6.2 IMPDH activities in late developmental stages of *ras* mutants ... 107
  - 3.3.6.3 Developmental northern analysis of wild-type strains ... 107
  - 3.3.6.4 The transcript pattern in the late developmental stages of *ras* mutants ... 111
  - 3.3.6.5 The transcript pattern in the early developmental stages of *ras* mutants ... 121

4 Discussion ... 127
- 4.1 IMPDH and the *ras* complex ... 127
  - 4.1.1 The 4.0kb *Eco* RI-*Hin* dIII fragment and the *ras* locus ... 127
  - 4.1.2 IMPDH and the *ras* complementation group ... 133
- 4.2 Possible models for regulation of the *ras* gene ... 136
- 4.3 *cis*-acting regulatory elements in *ras* gene ... 140
  - 4.3.1 *cis*-acting intronic splicing regulatory sequence ... 140
  - 4.3.2 *cis*-acting dosage compensating regulatory element ... 142
  - 4.3.3 Imperfect trinucleotide (CAG) repeat sequences ... 143
- 4.4 Oogenesis and the *cis*-acting intronic splicing enhancer ... 144
- 4.5 Conclusion ... 144
- 4.6 Future plans ... 145

Bibliography ... 148

Appendices ... 168

List of Figures

Figure 1-1 Model for the regulation of *Sxl* gene 21
Figure 1-2 The pathway of *de novo* IMP biosynthesis 24
Figure 1-3 Purine salvage pathway and interconversions 26
Figure 1-4 Purine catabolism 28
Figure 1-5 Structure of Drosophila *ade3* (GART) gene 32
Figure 1-6 Structure of pterin 36
Figure 2-1 Restriction map of λ EMBL-4 45
Figure 2-2 *Sau* 3A partially digested Drosophila genomic DNA 54
Figure 2-3 Preparation of a nested deletion series of cDNA SK83 58
Figure 2-4 Autoradiograph of a sample TLC plate 65
Figure 3-1 A genomic restriction map of the *ras* locus 69
Figure 3-2 Restriction maps of λ LF1 insert DNA and the homologous λ EMBL-4 clones from the wild-type (Am+) genomic library 70
Figure 3-3 *In situ* hybridization of the Am^{+} 7.3kb *Eco* RI fragment 71
Figure 3-4 Southern analysis of Am^{+} 3.3 and 4.0kb *Eco* RI fragments 72
Figure 3-5 The sequence at the 5′ end of the P-element insertion in λ LF1 73
Figure 3-6 Poly-A^{+} RNA probed by *ras* DNA fragments 75
Figure 3-7 Analysis of *ras* putative cDNAs digested by restriction enzymes with four base-pair target sites 76
Figure 3-8 Comparison of the nucleotide sequence of the 4.0kb *Eco* RI-*Hin* dIII fragment from λ SH4 and cDNA pSK29 79
Figure 3-9 Features in the "upstream" sequences of *ras* gene 83
Figure 3-10 "OPA" sequence hybridization pattern 85
Figure 3-11 Comparison of IMPDH sequences 87
Figure 3-12 Southern analysis of P-element induced lethal mutants 89
Figure 3-13 Southern analysis of Am^{+} and *ras* viable mutants 90
Figure 3-14 Sequence analysis of *ras* insert points 93
Figure 3-15 Diagram of the insertion sites of *ras*^{1}, *ras*^{2} and *ras*^{4} 94
Figure 3-16 Comparisons of the termini of the insertion sequence in the *ras*$^{1,\ 2\ \mathit{and}\ 4}$ mutants 95
Figure 3-17 Sequences of the 5′ LTR of the *ras*^{2} insert and the 3′ termini of the transposon blastopia polyprotein (BLPP) 97
Figure 3-18 Diagram of the sequence similarity between the *ras*^{2} insert and the BLPP 98
Figure 3-19 LTRs of *ras*^{4} and the copia-like transposon 101
Figure 3-20 37bp deletion in *ras*^{3} 102
Figure 3-21 Southern analysis of *ras* revertants 104
Figure 3-22 IMPDH activity of the wild-type females during later stages of development 105
Figure 3-23 IMPDH activity of wild-type females during later stages of development 106
Figure 3-24 IMPDH activity in late developmental stages 108
Figure 3-25 IMPDH activity in late developmental stages 109
Figure 3-26 IMPDH activities of young adult females 110
Figure 3-27 Northern analysis of 0-88 hours wild-type RNA 112
Figure 3-28 Northern analysis of wild-type RNA from late larval to young adult stage 113
Figure 3-29 Wild-type IMPDH RNA levels in 0-88 hour Drosophila embryos 114
Figure 3-30 Northern analysis of *ras*^{1} late larvae to young adult samples 117
Figure 3-31 Northern analysis of *ras*^{1} late larvae to young adult samples 118
Figure 3-32 Northern analysis of *ras*^{2} late larvae to young adult samples 119
Figure 3-33 Northern analysis of *ras*^{2} late larvae to young adult samples 120
Figure 3-34 Northern analysis of *ras*^{3} late larvae to young adult and *ras*^{4} young adult samples 121
Figure 3-35 Northern analysis of *ras*^{3}*m* late larvae to young adult and *ras*^{4} young adult samples 122
Figure 3-36 Northern analysis of 0-88 hour Am^{+} and *ras*^{2} samples 124
Figure 3-37 Northern analysis of 0-88 hour Am^{+} and *ras*^{2} samples 125
Figure 3-38 Northern analysis of 0-16 hour samples of OrR and *ras*$^{1-3}$ 126
Figure 4-1 Southern analysis of Am^{+} and *ras* viable mutants 131
Figure 4-2 Southern analysis of Am^{+} and *ras* viable mutants (Slee and Bownes) 132
Figure 4-3 Restriction map of *ras* locus (Slee and Bownes) 133

List of Tables

Table 1-1 Some well-known eukaryotic transcriptional regulatory proteins.. 11
Table 1-2 Purine biosynthesis involved enzymes and genes 25
Table 2-1 Description of *D. melanogaster* stocks 42
Table 2-2 Bacteria and the cloning vectors... 43
Table 2-3 Protocol for collection of embryos and early larvae... 49
Table 3-1 Tests of the genomic library cloning system.. 67
Table 3-2 The relationship of the cDNAs.. 78
Table 3-3 Similarity of the known IMPDH amino acid sequences 86
Table 3-4 Sequence comparison of internal segments of the *ras*^{2} insert with the BLPP sequence 99
Table 3-5 Developmental morphological change in the early embryos..127
Table 4-1 A summary of the expression of *ras-l* mutations in heterozygous combinations with *gua1*^{ts}, *pur1*^{1}, *pur1*^{2} and *ras*^{2}...140

List of Appendices

Appendix 1 Protein concentration of crude extracts 168
Appendix 2 IMPDH activity in late larval to young adult stages of wild-type (Am^+) strain 170
Appendix 3 IMPDH activity in the late developmental stage..........171
Appendix 4 Wild-type IMPDH RNA levels in 0-88 hour Drosophila embryos..........174
Appendix 5 Wild-type IMPDH RNA levels from late larval to young adult stage..........174
Appendix 6 Northern analysis of *ras*^{1} late larvae to young adult samples..........175
Appendix 7 Northern analysis of *ras*^{2} late larvae to young adult samples..........175
Appendix 8 Northern analysis of *ras*^{3} late larvae to young adult and *ras*^{4} young adult samples..........176
Appendix 9 Northern analysis of 0-88 hour Am^+ and *ras*^{2} samples..........176

1 Introduction

1.1 Gene regulation and purine metabolism

Metabolism, which includes all the chemical reactions in an organism, is a central theme for life. The process must be appropriately regulated for the organism to survive and reproduce. Most metabolic reactions are catalyzed by enzymes that are gene products. Thus a major component of this regulation is fulfilled by turning on and off the expression of enzyme encoding genes, thereby controlling and coordinating the gene activities harmoniously. In multicellular organisms, this regulation not only involves proper gene expression in the single cell, but also its coordination in space and time throughout development.

Molecular genetic analysis of well-defined biochemical pathways provides insights into gene regulation. The biochemical pathway of purine biosynthesis is well established; every reaction has been identified and all the enzymes involved are known (Henderson, 1972; Nash and Henderson, 1982; Patterson, 1985; Zalkin and Dixon, 1992). In bacteria and yeast, the genetics of purine biosynthesis has been studied extensively (Zalkin and Dixon, 1992). In animals, purines are some of the rather few fundamental or "primary" metabolites which can be synthesized endogenously. Therefore, understanding the molecular basis of the purine biosynthetic pathway may provide a source of unique insights for animal biology. In addition, the central metabolic importance of the purines themselves argues strongly for understanding of the system. Practically, the role of purines in nucleic acid synthesis suggests that purine biosynthesis is a natural target for chemotherapy. In consequence, purine metabolism has been investigated in numerous different organisms (Zalkin and Dixon, 1992).

Drosophila melanogaster purine mutants provide a suitable system to study the molecular biology of purine biosynthesis in multicellular eukaryotes. Since the 1960s, a number of purine mutants have been induced, recovered and analyzed (Vyse and Nash, 1969; Falk, 1973; Falk and Nash, 1974b; Naguib and Nash, 1976; Johnson *et al.*, 1977; Henikoff *et al.*, 1986b, c; Tiong and Nash, 1993; Clark, 1994; Nash *et al.*, 1994; Chomey and Nash, 1995; Clark, personal communication). The mutants have been used to investigate the genes encoding the enzymes in the pathway. Among these mutants, one interesting case has been called the *guanosine1-purine1-raspberry* (*gua^1-pur^1-ras*) complex by Nash *et al.* (1981). The complex yielded two different kinds of purine nucleoside auxotrophic mutants (Falk and Nash, 1974b), as well as several other mutant types. The locus is on the X chromosome. The mutants exhibit a complex complementation pattern. The first evidence that the complex might be involved in purine metabolism was the existence of the purine auxotrophs (Johnson *et al.*, 1979; Nash *et al.*, 1981); however, the locus was described much earlier on the basis of its prototype mutant, the dull red eye-colour *raspberry (ras,* 1-32.4, *9E3-4)*, which was first discovered by Muller in 1928 (Dunn, 1934; Muller, 1935; Beadle and Ephrussi, 1936). This thesis is focused on the molecular analysis of the locus.

1.2 Gene regulation

1.2.1 The natural history of gene expression

Genes are the fundamental functional units of heredity. In classical genetics, genes were considered as discrete physical particles present in all living organisms. Molecular genetics has revealed that genes are actually segments of DNA molecules. The genetic information is carried by the linear DNA polynucleotide sequences. During gene expression, functional RNA molecules and polypeptides are synthesized based on the genetic information. Genetic information can, somewhat arbitrarily, be divided into two sub-components, transcribed regions, whose information content acts as a template for RNA synthesis, and control regions, which bear the information that regulates transcription. This is an oversimplified distinction, for these two functions do overlap each other. Thus transcribed regions may have some regulatory functions and *vice versa*. Furthermore, it must be noted that some major aspects of regulation of gene expression are vested other than in the control of transcription.

A transcribed region acts as a template for the production (transcription) of an RNA molecule, the primary transcript. It extends from a 5′ transcription start-site to a 3′ termination site. Most transcribed regions include "coding sequences" which provide the information that is later translated into a polypeptide sequence. In bacteria, the coding region may specify several functionally related polypeptides adjacent to each other. These polypeptides are subsequently translated sequentially from the same messenger RNA (mRNA). Being located in the same transcript, they are subject to the same transcriptional control. This multi-functional transcription unit, the "operon", was first suggested by Jacob and Monod (1961). An mRNA transcribed from such an operon encodes several polypeptides and is said to be "polycistronic". In bacteria, the mRNA is usually the same as the primary transcript and translation commonly commences on the still growing (nascent) RNA molecule. The lac operon, the object of Jacob and Monod's work, remains the best known operon and will be discussed frequently where bacterial gene regulation is considered.

In eukaryotic organisms, the coding sequences normally encode only one polypeptide, that is, are "monocistronic". Transcription takes place in the nucleus and is uncoupled from translation, which occurs in the cytoplasm. Particularly in "higher" eukaryotes (plants and animals), the transcribed region is frequently longer than the mature RNA product, due to the presence of one or more intervening sequences that are not represented in the mature RNA (Gilbert, 1978; Dibb, 1993). During transcription, the entire region is transcribed into the primary transcript (heterogeneous nuclear RNA, hnRNA). Intervening sequences are then excised from the primary transcript and the coding sequences spliced together during RNA processing to generate mRNA which is used to direct synthesis of polypeptides. This process takes place within the nucleus and is coupled to transcription, so that spliceosomes (complexes in which processing occurs) form on the nascent RNA molecules (Padgett *et al.*, 1986; Green, 1991). The intervening DNA sequences are called "introns" and the sequences that comprise the mature RNA products are known as "exons" (Gilbert, 1978).

Gene expression is the entire process of decoding the encoded genetic information of a gene to produce a phenotypic effect. It includes transcription (RNA synthesis), processing (splicing and other post-transcriptional modifications of RNA), mRNA transportation (from nucleus to cytoplasm), translation (protein synthesis), post-translational modification of proteins. While gene regulation could mean the collective control at any step which can affect this process, the term is often reserved for control processes that affect transcription, processing and translation.

Of the various possible steps that might be subject to regulatory control, the first step, transcriptional initiation, is potentially the most important control point because it decides whether or not a gene will be transcribed and how much of the transcript will be produced.

1.2.2 Transcriptional regulation in bacteria

1.2.2.1 Regulation results from DNA/protein and protein/protein interactions

Transcriptional initiation is the result of the interaction between DNA regulatory elements located in the transcriptional control regions, RNA polymerase and transcriptional regulatory proteins. Transcriptional control regions are the DNA sequences which contain *cis*-acting regulatory sequences necessary for regulated transcription of a particular gene. These regulatory sequences are mostly binding sites for RNA polymerase and transcriptional regulatory proteins. Often, these sequences are outside the transcribed region and close to its 5′ end "upstream". However, both internal and downstream regulatory sequences are known, particularly in eukaryotes (see below). The interactions of transcription regulating proteins with each other, with specific regulatory DNA sequences and with RNA polymerase ultimately cause or impede the association of RNA polymerase with the gene and hence are fundamental to the regulation of gene expression.

1.2.2.2 Transcriptional control regions

DNase I footprinting (Dickson *et al.*, 1975; Schmitz and Galas, 1979) and electrophoretic mobility shift ("gel-shift") assay (EMSA, Fried and Crothers, 1981) experiments are major means by which the physical basis of transcriptional control has been examined. Much information has also been deduced from studies of regulatory mutants (McClure, 1985). Promoters, operators and activator binding sites are the most important DNA sequence motifs in the transcriptional control regions in most *Escherichia coli (E. coli)* operons (Collado-Vides *et al.*, 1991). Promoters are the binding sites for RNA polymerases. Operators are the binding sites of regulatory proteins called repressors, whose presence restricts transcriptional initiation. Activator sites are the binding sites of transcriptional activators (McClure, 1985; Gralla, 1991).

Most promoters are located upstream of the transcription start site. The majority of *E. coli* promoters are recognized by the σ initiation factor of polymerase (see below). There are two crucial DNA sequences in σ-sensitive promoters, a "-10 hexamer" (the Pribnow box) and a "-35 hexamer". The Pribnow box has the consensus sequence 5′-TATAAT-3′ and is located around 10 base pairs upstream of the transcription start site. The "-35 hexamer" has the consensus sequence 5′-TTGACA-3′ and is located (by the same convention) 35 base pairs upstream. In general, there is a strong relationship between the effectiveness ("strength") of a promoter and the degree to which these two elements agree with the consensus sequences (Siebenlist *et al.*, 1980; Hawley and McClure, 1983; McClure, 1985). Some *E. coli* operons may have more than one promoter (Goodrich and McClure, 1991). Newlands *et al.* (1992), Ross *et al.* (1993) and Rao *et al.* (1994) have reported an additional promoter component. A ~20bp AT rich sequence is located immediately upstream of the -35 hexamer in seven *E. coli* rrn (ribosomal RNA) genes. The rrn promoters are among the strongest promoters in *E. coli* and the rrn transcripts account for more than 60% of total RNA produced in a rapidly growing *E. coli* cell. The presence of this AT rich sequence increases the promoter efficiency at least 30-fold. Since this sequence is also bound by RNA polymerase, it is considered as a third component of an *E. coli* promoter and named upstream (UP) element.

The strength of a promoter seems to be a collective function of all the three elements. Strong promoters probably have all three elements, with the near-consensus Pribnow box and -35 hexamer plus UP elements. Weak promoters may have the Pribnow box plus either the -35 hexamer or a UP element, with near-consensus or, more likely, somewhat deviant sequences (Busby and Ebright, 1994).

Unlike promoters, operators may be located either upstream or downstream from the transcription start site. Most *E. coli* operator sequences are actually short inverted repeats. Each half of the inverted repeat is called a "half-site". This arrangement suggests that the repressors which bind the operators are probably protein dimers, with each monomer recognizing each half-site in the same way. The most extensively studied operator is the lac operator. The main lac operator (O_1) has imperfect symmetry of a 35bp inverted repeat with the center at position +11. Thus, each half-site of the lac operator forms a similar protein binding site but in the opposite orientation (Gilbert and Maxam, 1973; Schmitz and Galas, 1979). The function of the two halves is not equivalent with the upstream half having the higher binding affinity (Gilbert *et al.*, 1974, 1975; Horton *et al.*, 1997). There are two auxiliary operators in the lac operon. Operator O_2 is 402bp downstream from O_1 and O_3 is 93bp upstream from O_1. The functions of these three operators are not equal, the affinity of O_2 is ten-fold and O_3 is 300-fold lower than the affinity of main operator O_1 (Oehler *et al.*, 1990, 1994).

The lac operon is also subject to the *E. coli* catabolite activator protein (CAP). The CAP binding site is a 22 base pair imperfect inverted repeat with a conserved "TGTGA" motif in each half (de Crombrugghe *et al.*, 1984; Berg and von Hippel, 1988). Unlike promoters, CAP-sites are less well conserved, suggesting different affinity levels among CAP-sites, probably producing variant expression levels for different operons. CAP-sites are often located upstream from the transcription start site and on the same face of the DNA helix as the transcription start sites. The two preferred locations for CAP-sites are centered at -61/62 (as in the lac operon) and -41/42 (as in the gal operon). When the CAP-site is located at -41/42, it overlaps the promoter and replaces the -35 element (Ushida and Aiba, 1990; Zhou *et al.*, 1994).

Regulatory DNA motifs may be present with different combinations in different operons. In a typical example, the lac operon, there are, as mentioned above, at least six regulatory elements: From upstream, they are O_3 at about -82, CAP-site at about -61/62, -35 hexamer, -10 hexamer (Pribnow box), O_1 at about +11 and O_2 at about +412.

1.2.2.3 Regulation of transcriptional initiation

The *E. coli* RNA polymerase holoenzyme (RNAP) consists of two elements, one specific σ initiation factor, which mediates promoter specificity, and the core enzyme, which contains all the functions necessary for RNA synthesis. The core enzyme has four subunits, β' (165 kDa), β (155 kDa) and two identical α subunits (35 kDa). One of the σ factors, σ^{70} recognizes many *E. coli* promoters and binds the -10 and -35 elements. σ^{70} is a 613 amino acid polypeptide with four conserved amino acid sequence motifs common to all σ factors. σ factors are required for the core enzyme to access the promoter and initiate transcription precisely. After initiation, σ factors dissociate from the core enzyme, which continues the transcription (Gross *et al.*, 1992; Keener and Nomura, 1993; Eick *et al.*, 1994).

The two large subunits of core enzyme, β' and β, contact the DNA template directly and carry on the RNA synthesis. The two small α units each contain two functional domains, the N-terminal domain and C-terminal domain. The N-terminal domain plays an important role in polymerase assembly. The N-terminal domain is necessary and sufficient to initiate the assembly pathway $\alpha \rightarrow \alpha_2 \rightarrow \alpha_2\beta \rightarrow \alpha_2\beta\beta' \rightarrow \alpha_2\beta\beta'\sigma^{70}$. The C-terminal domain makes contact with DNA motifs (for example, the UP elements) or regulatory proteins (activators). The C-terminal domain works like an internal activator; without it, the polymerase can still initiate transcription but only at the basal level (Eick *et al.*, 1994; Busby and Ebright, 1994; Czarniecki *et al.*, 1997; Murakami *et al.*, 1997).

Repressors and activators, the two most important classes of regulatory proteins in prokaryotic cells, act at the level of transcriptional initiation. Most are homodimers or homotetramers with two or more functional domains. They make contacts with DNA motifs, other regulatory proteins, RNA polymerases and/or their substrates or other molecules with affinity for regulatory proteins (ligands), in fulfilling their regulatory functions. Since repressors and activators contain similar physical structures, it is not uncommon that a regulatory protein has dual functions, working as either a repressor or an activator. This functional switch occurs when the regulatory proteins bind different promoter regions or are under different cellular contexts, such as the λ cI protein (Ptashne, 1992; Li *et al.*, 1997), CAP and the catabolite repressor-activator (Cra) protein (Ramseier, 1996; Saier and Ramseier, 1996). Some regulatory proteins function on more than one operon. The set of operons regulated by one regulatory protein has been called a regulon (Jones and Aizawa, 1991; Newman *et al.*, 1992).

In the same way that the prototype operator, the lac operator, is very well understood, so is the lac repressor, its binding protein. The *E. coli* lac repressor (LacR) is a homotetramer which consists of two dyad-symmetric dimers. It has four DNA binding domains (the N-terminal domains from each of the four monomers) on the same side of the tetramer (Friedman *et al.*, 1995; Lewis *et al.*, 1996; Matthews, 1996). Each DNA binding domain can bind one operator half-site, so the LacR tetramer can potentially bind two operators at the same time. The C-termini of the monomers form the tetrameric core which binds lactose. Under repression, one of two dimers binds the main operator O_1 and the other dimer binds one of the two auxiliary operators O_2 or O_3. This binding bends DNA between the two operators and forms a DNA repression loop which prevents transcriptional initiation. Since O_1 overlaps the lac promoter, the physical binding of O_1 by LacR also interferes with the polymerase binding and further impedes initiation (Flashner and Gralla, 1988; Gralla, 1989; Friedman *et al.*, 1995; Lewis *et al.*, 1996).

Activators facilitate transcriptional initiation when they bind to the binding sites and make contact with RNA polymerase. The best understood bacterial activator is the *E. coli* CAP, also referred to as cAMP (cyclic AMP) receptor protein (CRP). CAP is a dimer with two identical subunits. The N-terminal domain of the subunits can bind cAMP and the C-terminal domain contains motifs with affinity for both DNA and protein (McKay *et al.*, 1982; Weber and Steitz, 1987). Each of the DNA binding motifs can bind one half-site of the CAP-site (Zhou *et al.*, 1993) and each of their protein binding motifs, or "activating regions", can bind one of the α C-terminal domains of the polymerase (Niu *et al.*, 1994; Tang *et al.*, 1994). When cAMP binds CAP, the cAMP-CAP complex has high affinity for the CAP-site. Thus formed the CAP-DNA complex bends the DNA and allows CAP direct contact with the α C-terminal domain, which activates transcription (Busby and Ebright, 1994; Bokal *et al.*, 1997).

In *E. coli*, glucose is the preferred sugar source. When both glucose and lactose are present, glucose is used first, and lactose is used only after glucose has been depleted. When lactose is not being used, either because it is not present or in the presence of glucose, lactose metabolism is at a very low level. However, when lactose is used, the lactose metabolism level increases up to 1,000 times in a short period. In this case, the transcription level of the lac operon increases before the lactose metabolism (Jacob and Monod, 1961; Yanofsky, 1992). In addition to lactose, lac operon transcription is also regulated by cAMP, which is a product associated with glucose depletion (Jacob and Monod, 1961; Zubay *et al.*, 1970). When lactose is present, it binds to the repressor, which changes conformation and loses its affinity for the operator. This switches the lac operon to a configuration that favours transcription and could promote lactose metabolism. If at the same time *E. coli* is starving for glucose, the cAMP level increases and so does the cAMP-CAP complex which binds the CAP-site, resulting the activation of lac operon. When lactose is absent, the lac repressor binds the operator and turns it off.

A further interesting refinement of lac operon regulation is the fact that the cAMP-CAP complex can also enhance repression of lac transcription. The binding of CAP to the CAP-site facilitates lactose-free lac repressor binding to O_1 and O_3. CAP and lac repressor can bind simultaneously on the lac promoter and this cooperation forms a more stable repressor-operator complex (Hudson and Fried, 1990). Vossen *et al.* (1996) have observed that the CAP binding stabilizes the repressor-operator complex approximately 11-fold. Thus, when both glucose and lactose are absent, increased cAMP ensures that the lac repressor binds stably to the promoter, saving precious energy, or utilizing it in exploiting some alternative energy source (Lewis *et al.*, 1996).

The situation in the *E. coli* lac operon is found in variant form throughout the bacteria. Typically, transcriptional initiation is a crucial and, in many operons, the only controlling step in regulation of gene expression. The transcriptional regulation of many prokaryotic operons is intimately tied to the core promoter and the function of RNA polymerase. RNA polymerase holoenzyme (ααββ′σ) is sufficient to initiate a basal (unregulated) level of transcription from the promoter. The interactions between the regulatory DNA sequences, regulatory proteins and polymerase decide the actual transcription level. The repression/activation switching system is elegant and simple. It provides an appropriate conceptual starting-point for an examination of regulation in eukaryotes.

1.2.3 Transcriptional regulation in eukaryotes

1.2.3.1 Prokaryotic and eukaryotic transcriptional systems

Prokaryotic and eukaryotic transcriptional systems share many features. The prokaryotic transcriptional system can be considered as a simplified analogue of the much more complex eukaryotic system. However, there are major differences between these two systems (Eick *et al.*, 1994; Goodrich *et al.*, 1996; Ranish and Hahn, 1996; Roeder, 1996).

Unlike prokaryotic DNA, eukaryotic DNA is normally complex with histones and other proteins in the form of chromatin. Histones by themselves generally prevent polymerase access to DNA templates. In vertebrates, the cytosine in 5′-CpG-3′ dinucleotides is frequently methylated, which is also known to repress transcription (Siegfried and Cedar, 1997). Considering the huge number of genes in eukaryotes, especially in vertebrates, any unregulated gene expression can be fatal to the organism. Histones, along with methylation, can efficiently repress spurious gene transcription (Bird, 1995; Klein and Costa, 1997). So, essentially all eukaryotic genes are potentially transcriptionally inert. Thus specific activation instead of repression plays a key role in eukaryotic gene expression.

In multi-cellular systems, cells develop and differentiate into different tissues to fulfill different functions. Thus gene expression in each individual cell is regulated differentially in order to meet the requirements of the whole organism, not just the cell itself. In addition, with minor exceptions, eukaryotic DNA is restricted to the nucleus and separated from cytoplasm by the nuclear membrane. As a result of these two layers of complexity, intercellular interactions mediated by signal transduction systems as well as intracellular communication across the nuclear membrane, become major features of gene regulation. The collective functioning of these systems regulates gene expression temporally and spatially during the individual life span.

The eukaryotic transcriptional apparatus is also more complex than that in prokaryotes. There are three different RNA polymerases in the eukaryotic nucleus, RNA polymerase I, II and III. Polymerase I synthesizes rRNA (ribosomal RNA). Polymerase II synthesizes mRNA. Polymerase III synthesizes tRNA (transfer RNA) and other small RNAs (Roeder, 1976; Sentenac *et al.*, 1992). Each polymerase interacts with a different set of regulatory proteins and transcribes a distinguishable set of genes with different promoter structures. This multiplicity of polymerase forms and specificities can be regarded as paralleling the *E. coli* polymerase holoenzyme with its several different σ factors.

The basic differences in lifestyles between prokaryotes and eukaryotes force them to choose divergent strategies to regulate their gene expression. In most bacteria, gene regulation seems to be focused on rapidly adjusting their metabolism and behavior to survive in the changing environment. In eukaryotic cells, gene regulation is more focused on the control of development, differentiation and the cell cycle (Yanofsky, 1992).

1.2.3.2 The basics of eukaryotic transcriptional initiation

Of the three major eukaryotic RNA polymerases, polymerase II plays a predominant role because it transcribes majority of eukaryotic genes, the protein coding genes. The transcriptional control regions of polymerase II transcribed genes are composed of two functional sub-regions, the core-promoter and the gene-specific DNA regulatory elements (Roeder, 1996). The core-promoter consists of the minimal DNA elements necessary for sufficient and accurate transcriptional initiation and located around the transcription start site. The gene-specific DNA regulatory elements, also referred to variously as enhancers, silencers, promoter-proximal elements, upstream activation sequence (UAS) or regulatory regions (McKnight, 1982; Dynan, 1989; Roeder, 1991; Kornberg, 1996), are diverse in nature. They have, in common, the ability to activate or repress transcriptional initiation in a gene and situation-specific manner.

There are two common eukaryotic core-promoter elements, the TATA box and initiator (Inr). They can function independently or synergistically. There are three major groups of core-promoters, $\text{TATA}^{+}\text{Inr}^{+}$, $\text{TATA}^{+}\text{Inr}^{-}$ and $\text{TATA}^{-}\text{Inr}^{+}$ (Goodrich *et al.*, 1996; Roeder, 1996). The TATA box (consensus $\text{TATA}^{A}/_{A}\text{A}$), which is the primary target for the control of initiation, is similar to the Pribnow box of bacterial promoters and located about 25-30bp upstream of the transcription start site (Breathnach and Chambon, 1981). Inr encompasses the transcription start site and was initially identified in the TATA^{-} promoter in the terminal deoxynucleotidyl transferase gene (TdT) by Smale and Baltimore (1989). It is a short pyrimidine-rich sequence with a highly conserved dinucleotide, C at position -1 and A at +1 relative to the transcription start site (Weis and Reinberg 1992; Javahery *et al.*, 1994). Novina and Roy (1996) have suggested there is a fourth type of core-promoter, $\text{TATA}^{-}\text{Inr}^{-}$, with multiple transcriptional start sites.

The core-promoter is recognized by a universal transcriptional machinery consisting of RNA polymerase II and a group of accessory proteins named general transcription initiation factors. This universal machinery, referred to as the basal apparatus, has an intrinsic ability to initiate low levels of accurate transcription from core promoters *in vitro*. The level of transcription defined in this manner is referred to as basal transcription. The subunits of the basal apparatus are conserved from yeast to man (Kornberg, 1996).

Eukaryotic RNA polymerase II contains 12 evolutionarily conserved subunits. The two largest subunits, B220 and B150 in yeast, share significant sequence similarity with the *E. coli* core enzyme β' and β subunits, respectively, and two other identical subunits, B44 in yeast, are homologous with the α subunit. These four subunits form the core of the polymerases and have homologues in all three eukaryotic polymerases (Allison *et al.*, 1985; Sawadogo and Sentenac, 1990, Young, 1991; Sentenac *et al.*, 1992; Ito and Nakamura, 1996).

Polymerase II can neither initiate transcription by itself even from the strongest promoters (Weil *et al.*, 1979), nor elongate nascent RNA (Reines *et al.*, 1996). Prior to transcription, general transcriptional initiation factors assemble at the promoter region along with the polymerase into a pre-initiation complex (PIC). The general transcription initiation factors, as a group, play a role similar to the prokaryotic σ

factors (Zawel and Reinberg, 1995). There are six well-known polymerase II associated general initiation factors, TFIIA, B, D, E, F and H. Among them, TFIID is the only one with the capacity for sequence-specific binding to core-promoters. TFIID is actually a multifunctional promoter-recognition complex which contains at least ten subunits; one of them, TATA box-binding protein (TBP), directly recognizes and binds the TATA-box. The remaining TFIID subunits are collectively named TBP-associated factors (TAFs) and are known to be directly or indirectly involved in PIC assembly, especially in $TATA^{-}Inr^{+}$ core-promoters.

The *in vitro* assembly of a minimal PIC on a strong promoter, like the adenovirus major late promoter (AdML, $TATA^{+}$ Inr^{+}), involves the following steps: First, TBP, the subunit of TFIID, binds to the TATA box and forms a stable complex. Then, TFIIB binds to this complex by direct interactions with TBP and the surrounding DNA sequences. Next, a pre-formed TFIIF-polymerase complex directly contacts and binds to TFIIB of the DNA-TBP-TFIIB complex. Finally, TFIIE and TFIIH bind, directly contacting polymerase II. Transcriptional initiation can now take place (Burley and Roeder, 1996; Roeder, 1996; Nikolov and Burley, 1997). TAFs and TFIIA are not essential for the assembly of a minimal PIC on a strong promoter, like AdML. They do sometimes have important roles in PIC assembly, especially on TATA-less promoters (see below). However, their presence stabilizes the PIC *in vitro*. Thus, intact TFIID is more efficient than its purified TBP subunit in core-promoter recognition. TFIIA can bind the TFIID and the promoter directly and form a TFIID-TFIIA-promoter complex which is more stable than the TBP-promoter complex (Zawel and Reinberg, 1993; Buratowski *et al.*, 1989; Geiger *et al.*, 1996; Tan *et al.*, 1996).

In vitro experiments reveal that TFIID, probably in company with TFIIA, disassociates from the basal apparatus after transcriptional initiation. They remain stably associated with the promoter and ready for the next round of PIC assembly. Other general initiation factors, TFIIB, TFIIE and TFIIH also drop off the transcription machinery after the initiation, leaving polymerase-TFIIF to continue the elongation (Roeder, 1996).

In vitro transcription from the murine TdT promoter ($TATA^{-}$ Inr^{+}), a TATA-less promoter, reveals that PIC assembly cannot take place at all unless all the general initiation factors are present, including TFIIA and the TAFs. However, it appears that TBP is not the primary factor for PIC formation on this TATA-less promoter. Two TAFs, $TAF_{II}250$ and $TAF_{II}150$, seem to play key roles in the $TATA^{-}Inr^{+}$ promoter recognition (Kaufmann and Smale, 1994; Hansen and Tjian, 1995; Verrijzer *et al.*, 1995). TFIIA seems to be required for optimizing the function of TAFs (Zawel and Reinberg, 1993; Ma *et al.*, 1996).

Polymerase II promoters are extremely diverse. As mentioned above, the strength of a core-promoter depends on the elements involved and their arrangement. Furthermore, the specific DNA sequences in the elements and the distance between them may vary greatly from promoter to promoter. These differences presumably condition different basal levels of transcriptional activity and set the stage for regulation of even more complex patterns of variable activity for the individual genes (Smale *et al.*, 1990; Hernandez, 1993; Sadowski *et al.*, 1993; Henry *et al.*, 1995; Ruppert *et al.*, 1996).

1.2.3.3 Transcriptional activation in eukaryotes

In vivo, the eukaryotic genes are bound by histones, which by themselves generate an over-riding constraint upon gene activity. Although the general transcription machinery has an intrinsic ability to initiate basal transcription at core-promoter elements *in vitro*, there is little evidence that it can initiate RNA synthesis in chromatin *in vivo*. Batteries of regulatory proteins and their binding sites, the DNA regulatory elements, play the crucial role in mediating transcriptional initiation *in vivo* (Goodrich *et al.*, 1996; Kaiser and Meisterernst, 1996; Verrijzer and Tjian, 1996).

Transcriptional regulatory elements are DNA sequences which dramatically affect transcription rates either positively or negatively (Serfling, *et al.* 1985; Gotta and Gasser 1996). In this sense, regulatory elements are, of course, also present in bacteria (Kustu *et al.*, 1991; Ranish and Hahn, 1996). The regulatory elements are recognized and bound by regulatory proteins, which in turn modulate the initiation of transcription. The regulatory elements can be located near the transcription start site or at a considerable distance upstream or downstream of a gene, or within a gene. Those which are at a significant distance from the initiation site can generally function in either orientation with respect to the core-promoter (Maniatis *et al.*, 1987). This property is probably a trivial one, simply depending upon the ability of DNA to twist and form loop structures, allowing establishment of an identical configuration at the critical site for regulation. Regulatory elements normally contain one or more discrete protein binding sequence motifs (Dynan, 1989; Tjian and Maniatis, 1994). The different content and arrangement of these DNA motifs generates possibilities for unique protein assemblies, each with its own regulatory properties. The repertoire of active regulatory proteins differs between cell types. Thus, the specific interaction between regulatory proteins and DNA regulatory elements results in cell specific gene expression (Serfling, *et al.* 1985; Muller, *et al.* 1988; Levine and Hoey 1988; Bonifer, *et al.* 1996).

Table 1-1 lists some well-known regulatory proteins and their target DNA sequences and is provided for reference: these systems will be discussed further in the subsequent text.

Some regulatory proteins operate simply, by binding to and bending DNA, thereby allowing other regulatory proteins access to the DNA. More typically, transcription-regulating proteins contain a DNA-binding domain and one or more transcription-regulating domains. The DNA-binding domain binds the site-specific DNA motifs. The transcriptional activation domains, either directly or indirectly, target the general transcriptional machinery and/or other regulatory proteins.

Many regulatory proteins also have multimerization domains which allow the formation of homo- or hetero-multimers. Hetero-multimer formation between related regulatory proteins expands the repertoire of regulatory protein functions, particularly allowing interaction with more DNA motifs, for example. On the other hand, the formation of hetero-multimers may also serve to neutralize the regulatory function of one or more of the component polypeptides (Roeder, 1991; Tjian and Maniatis, 1994).

Table 1-1 Some well-known eukaryotic transcriptional regulatory proteins

Symbol	Name	Recognizing motif
ATF family	Activating transcription factor	GTGACGT$^{AA}{}_{CG}$
C/EBP	CCAAT/Enhancer binding protein	CCAAT
E1A, E2F	Viral immediate-early proteins	TATA box
GAL4	Yeast transcriptional activator protein	CGGAGGACTGTCCTCCG
GATA-1	GATA-1	TGATAG
GCN4	Yeast transcriptional activator protein	ATGACTCAT
Sp1	Specificity protein 1	GGGGCGGGGC

1.2.4 Regulation of house-keeping and tissue-specific genes

1.2.4.1 Structure and regulation of house-keeping genes

House-keeping genes encode proteins which carry on the basic functions common to all cells. They are expressed in almost every cell type. Tissue-specific genes encode proteins with special functions and are expressed in selected cell types and/or at certain developmental stages. The differences between these two kinds gene are vague, because, ultimately, any gene has the potential to be expressed ubiquitously, while so-called house-keeping genes can clearly be regulated in a tissue-specific manner under certain circumstances.

At the DNA sequence level, a distinction has been made between promoters associated with the two classes of genes in mammalian cells (Dynan, 1986). House-keeping genes commonly do not appear to have TATA-box and CCAAT elements in their regulatory regions. Often, their regulatory regions contain initiators and very GC rich sequences (CpG islands). Multiple regulatory protein binding sites are present in this GC rich sequence, especially Sp1 binding sites. The rat type I hexokinase gene is typical since it does not have a TATA-box. There is an approximately 1kb CG rich sequence in the regulatory region. The CG rich sequence contains three initiation start sites located about 460, 300 and 100bp upstream from the translation start codon (White *et al.*, 1996).

Near ubiquitous expression of house-keeping genes does not imply lack of regulation (Dynan, 1986). In CHOC 400 cells, dihydrofolate reductase (DHFR) expression increases about four times during G1-to-S phase of the cell cycle. Gel mobility shift assays have shown that two different species of E2F like proteins interact with the DHFR promoter during the cell cycle. The first E2F like protein prefers the site TTTGGCGC and its binding activity is generally constant during the cell cycle. The second one prefers to bind the site TTTCGCGC and its binding activity increases during the G1-to-S period. These results indicate that the first E2F-like protein probably participates in basal transcriptional initiation and the second E2F like protein may be involved in the regulation of gene expression during G1-to-S phase transition (Wells *et al.*, 1996).

House-keeping genes can also be tissue-specifically regulated. The hypoxanthine phosphoribosyltransferase (HPRT) gene is constitutively expressed at low levels in all mammalian tissues, but at notably higher levels in the brain. Rincon-Limas *et al.* (1994, 1995) have reported that a 182bp element (HPRT-NE) within the 5′ flanking region plays a double role in HPRT regulation, activating HPRT expression in neuronal cells and holding HPRT expression at the basal levels in non-neuronal tissues. In neuronal cells, HPRT-NE interacts with a group of neuronal-specific nuclear proteins (complex I) and activates the gene expression. The genes encoding the nuclear proteins of complex I are expressed following neuronal differentiation, which correlates with an increase of HPRT gene transcription. In non-neuronal tissues, HPRT-NE interacts with another group of nuclear proteins (complex II) which acts as a repressor (Jiralerspong and Patel, 1996).

As mentioned above, the differences between house-keeping and tissue-specific genes are ultimately vague. In some cases, the same gene can have two distinct functions, one of which can be classified as house-keeping and the other as tissue-specific. In the duck, lactate dehydrogenase B (LDH-B) is a house-keeping enzyme in most cells, but acts as a lens structural protein (ε-crystallin) in the eye. It is encoded by a single gene, LDH-B/epsilon-crystallin (Stapel *et al.*, 1985; Wistow *et al.*, 1987; Hendriks *et al.*, 1988). This must be the result of recruitment of a house-keeping gene to a tissue-specific function during evolution.

The duck LDH-B/ε-crystallin gene has the typical house-keeping gene structure. It contains two closely spaced transcription start sites, an initiator at -28 and an HIP1 sequence at +1, and a CpG island. Transcription is predominantly initiated from the +1 site in the lens but both initiation sites are used equally in the heart (Kraft *et al.*, 1993). The region -70/+18 is required for basal level transcription and it also enhances promoter activity in the lens. There is a critical Sp1 binding site at -50. When this site is mutated to a non-Sp1 binding sequence (CCGCCC to CTTTCC), preferential expression in the lens is abolished. Since Sp1 is a ubiquitous protein, it seems likely that there is a Sp1-like protein with special properties which binds to this site in the lens.

Another important regulatory region is from -22 to +2, which is bound by a protein present in heart extracts but not in lens extracts. It seems there is a nuclear regulatory factor present in heart but not in lens. The authors present a model in which a factor transmitting the effect of Sp1 or Sp1-like factor, bound at -50, to the transcription initiation complex is responsible for preferential expression in the lens, an effect which is eliminated by the binding of a factor at -22/+2 in the heart (Brunekreef *et al.*, 1996).

1.2.4.2 Tissue-specific gene expression

Tissue-specific regulation is achieved by a variety of mechanisms, such as use of alternative promoters, alternative splicing (see next section), unmethylation or changes of hormone level, working alone or in combination. The regulatory region of tissue-specific genes often contains a TATA$^+$ core-promoter with or without the initiator element. There are often multiple protein binding domains in the regulatory region. These DNA domains interact with many tissue-specific regulatory proteins to control specific gene expression.

The rat preprotachykinin-A (rPPT) gene is expressed in a subset of dorsal root ganglion neurons. A region of about 1,300bp upstream of the transcription start site has been extensively analyzed (Quinn *et al.*, 1995). There are multiple regulatory domains in this region. The interactions between these DNA domains and a variety of tissue-specific, developmental, and stimulus-induced regulatory proteins are involved in the regulation of gene expression. The regulatory domains are required for transcriptional activation in many neuronal cells but are required to repress transcription in non-neuronal cells. When the domains are removed or modified, gene expression occurs in non-neuronal cells. The results suggest that rPPT gene tissue-specific expression is controlled by the dynamic interplay of both positive and negative regulatory elements in different cells.

Alternative promoters are often used in tissue-specific regulation. The human CYP19 gene encodes aromatase cytochrome P450 (P450arim). Several promoters are present in its regulatory sequences and are used in different tissues. The transcripts from these promoters contain unique 5′-noncoding termini.

The proximal promoter is used in gonads and primarily regulated by cAMP. The distal promoter, which is located 40kb upstream from the proximal promoter, is used in placenta and is regulated by retinoids (Simpson *et al.*, 1997). Most transcripts in adipose tissues are transcribed from a third promoter. The expression in breast adipose tissue is positively related to estrogen levels (Simpson *et al.*, 1994; Bulun and Simpson, 1994). A common 3′-splice site upstream from the translation start site is used in all tissues. Thus even though the mRNAs are transcribed from different promoters in different tissues, all of them contain an identical coding sequence and produce an identical polypeptide (Toda *et al.*, 1996; Agarwal *et al.*, 1997; Simpson *et al.*, 1997).

House-keeping genes can be temporally and spatially regulated as tissue-specific genes during development by using alternative promoters. The murine GAD67 gene encodes the 67-kD form of glutamic acid decarboxylase. It contains three promoters. The promoter P1 has typical housekeeping promoter features (Sp1 binding site and CpG island) and corresponds to major transcripts in most tissues. Promoter P2 and P3, 130 and 295bp upstream from the major transcription start site, contain TATA and CAAT-like elements. There are multiple regulatory protein binding sites in the 5′ upstream region. These regulatory DNA elements may play important roles in developmental and tissue-specific expression (Szabo *et al.* 1996).

Another example of the use of alternative promoters is human α folate receptor (α hFR) gene. It contains seven exons and the polypeptide is encoded within exons 4 to 7. Sequencing revealed that there are two promoters in the α hFR gene, one upstream of exon 1 (P1) and one upstream of exon 4 (P4). P1 is mostly used in cerebellum and kidney and P2 is mostly used in lung and KB cells (a cultured cell line derived originally from a human epidermoid carcinoma). Seven different cDNAs have been reported (Elwood *et al.*, 1997). All share the same ORF but have unique 5′-termini. It is the alternative transcription from two promoters and alternative splicing that generate the tissue-specific alternative mRNA expression. Alternative splicing, which will be discussed in the next section, adds additional complexity in the regulation.

As mentioned before, methylation is often found on the C of CpG dinucleotides in vertebrate DNA. Actively transcribed genes generally lack methylated C in their regulatory sequences. Thus, the absence and presence of methylation often distinguish between the activity and inactivity of genes in cases of tissue-specific regulation. The murine Pdha-2 gene codes for a male spermatogenesis-specific isoform of the E1a subunit of pyruvate dehydrogenase. Transgenic experiments show that a minimal promoter of 187bp from the gene is sufficient to initiate the tissue-specific expression of a CAT reporter gene (Iannello *et al.*, 1997). Further experiments reveal that methylation of CpG dinucleotides in the minimal promoter represses expression both *in vitro* and *in vivo*. In this case, the absence of testis-specific methylation of the Pdha-2 permits expression in spermatogenesis.

Although a TATA-box is frequently present in tissue-specific genes, it is not invariably needed for tissue-specific regulation. Some tissue-specific genes have a typical house-keeping promoter structure. The human platelet/endothelial cell adhesion molecule-1 gene is expressed in platelets, endothelial cells and certain leukocyte subsets. It does not have a TATA-box or CAAT binding sites in its regulatory region. Transcription is initiated at several closely spaced sites in the initiator sequences. Several regulatory protein binding sites, including Sp1 and the GATA element, are located in the 300bp region upstream from the initiator (Gumina *et al.*, 1997). There is no TATA-box in the rat luteinizing hormone receptor gene either (Dufau *et al.*, 1995). The promoter contains an initiator element and two Sp1 domains. Sp1 protein binding is required for transcriptional initiation. Tissue-specific expression is achieved by multiple trans-regulatory proteins binding to different DNA regulatory elements.

In summary, then, transcription of most eukaryotic genes is regulated by the interaction between multiple *cis*-acting transcription sequences and multiple *trans*-acting DNA-binding proteins. The interactions enable RNA polymerase to recognize and bind to the promoter and to commence transcription accurately. In more complex cases, the activity of different promoters may be evoked to produce specific patterns of gene expression. As we shall see later, the process of RNA processing, using alternative splicing patterns, can also lead to complex differences in gene expression between tissues.

1.3 Introns

1.3.1 Intron structure

Introns were discovered in 1977 in several laboratories (Berget *et al.*, 1977; Brack and Tonegawa, 1977; Breathnach *et al.*, 1977; Chow *et al.*, 1977) and named by Gilbert (1978) for "intragenic regions". Since then, the intron, which may be inherited from the primitive genes (Blake, 1978, 1983; Doolittle, 1978; Gilbert, 1978, 1979) or acquired during evolution (Cavalier-Smith, 1978, 1991; Rogers, 1985; Rogers, 1989), has been recognized as a frequent element in eukaryotic gene organization.

Introns are often much longer than exons, the elements whose transcripts are ultimately joined together to form the mRNA. Intronic, as well as exonic sequences, are represented in primary transcripts. In vertebrates, the length of most introns ranges from 10^3 to 10^5bp, although many are smaller than 10^3bp. However, Drosophila introns are much smaller, with the average of 79bp, which is extremely small compared to mammalian introns (Mount *et al.*, 1992). In contrast, exons are normally 300bp or less, with an average about 130-140bp (Berget, 1995).

The intron/exon structure is often closely related to the structure of the encoded polypeptide. Gilbert (1978, 1979) suggested exons to be equivalent to "protein modules", a contiguous polypeptide segment which is assumed to have compact molecular configuration. Mitiko Go (1981) developed an analytical tool (the "Go plot") to study the correlation between gene structure and protein structure and successfully predicted intron position based on protein information (Jensen *et al.*, 1981; Moens *et al.*, 1992). By extension, it was suggested that the exonic translation products represent primitive elements of protein structures (Gilbert, 1978; Go, 1981).

The DNA sequence structure of introns has been extensively studied (Smith *et al.*, 1989; Green, 1991). There are four essential elements related to the manner in which a transcript is converted to mRNA (pre-mRNA splicing): the 5′ and 3′ splice sites, the branchpoint and a pyrimidine-rich region between the branchpoint and 3′ splice site. Splicing of pre-mRNA takes place *via* two transesterification reactions. First the 5′ splice site is hydrolyzed and the 5′ end of the intron concomitantly joins the A of the branchpoint through a 2′-5′ phosphodiester bond, forming a branched or "lariat" intermediate. Then, the 3′ splice site is hydrolyzed and the two exons join, leaving the intron as a lariat structure.

The sequences of these splice determining elements of the pre-mRNA are short and relatively poorly conserved. They provide little information for splice site selection, especially for where alternative and/or weak splicing signals are available (Rio, 1993). Additional information carried by "splicing enhancers" is required for splicing site selection (Tian and Maniatis, 1992; Watakabe *et al.*, 1993). Splicing enhancers are RNA sequences bound by proteins and often contain multiple repeats and/or purine-rich motifs (Hertel *et al.*, 1996). They share many features with transcriptional regulatory elements and function in a developmental and tissue-specific manner (Hertel *et al.*, 1997). The majority of splicing enhancers are located in exons, but some of them are within introns (Carlo *et al.*, 1996; Gallego *et al.*, 1997; Min *et al.*, 1997). Exonic splicing enhancers normally assist the previous (upstream) 3′ splice site selection. However, some enhancers can also help the recognition of 5′ (the next, downstream) splicing sites and the branchpoint region (Tian and Maniatis, 1992; Chabot, 1996).

1.3.2 Spliceosome assembly

The U-rich small nuclear RNAs (snRNAs, including U1, U2, U4, U5 and U6), and many different proteins, are involved in the splicing reactions. The snRNAs and proteins assemble into a large and dynamic ribonucleoprotein complex, the spliceosome, in an ordered series of events (Moore *et al.*, 1993). The five snRNAs are the RNA subunits of four small nuclear ribonucleoproteins (snRNPs), U1, U2, U5 and U4/U6, which are the main components of the spliceosome.

Certain snRNAs contain short, conserved sequences complementary to the splice sites and the branchpoint region of the pre-mRNA (Moore *et al.*, 1993; Horowitz and Krainer, 1994). The 5′ end of U1 is complementary to both the 5′ and 3′ splice sites. The base-pairing between 5′ splice site and U1 RNA is necessary for use of the splice site. U2 contains sequences complementary to the branchpoint region; when U2 base-pairs with branchpoint sequences, the branch site A is excluded from the pairing. U5 interacts with exons adjacent to both splice sites; this interaction probably causes a RNA loop formation and brings the two exons together.

There are also several cases of complementary sequences which allow different snRNAs to base pair. U4 and U6 have segments which contain extensive complementary sequences and can base-pair to each other, forming a U4/U6 complex. Another segment of U6, which is adjacent to the U4/U6 complementary sequence, contains sequences complementary with U2 and can form a U2/U6 helix by base-pairing. U6 can also base-pair with the 5′ splice site. The U2/U6 interaction probably brings the 5′ splice site and the branchpoint together. In general, the pre-mRNA-snRNP and snRNP-snRNP interactions generate a multi-molecular architecture that promotes pre-mRNA splicing, in a manner that can be regarded as catalysis (Madhani and Guthrie, 1994; Nilsen, 1994).

Many proteins are involved in pre-mRNA splicing (Dreyfuss *et al.*, 1993; Wu and Maniatis, 1993; Will and Lührmann, 1997). One group of them is the SR protein (serine and arginine rich) family which plays an essential role in spliceosome assembly (Zahler *et al.*, 1992; Valcárcel and Green, 1996; Mount, 1997). All SR proteins contain an N-terminal RNA-binding domain composed of one or two ribonucleoprotein consensus motifs (RNP-CS) and a C-terminal arginine and serine rich domain (RS domain) composed of multiple arginine-serine dipeptides. RS domains of different SR proteins can interact with each other (Wu and Maniatis, 1993; Kohtz *et al.*, 1994) and some of them can also promote RNA-RNA pairing (Lee *et al.*, 1993). The structure of SR proteins suggests that they recognize specific RNA splicing sequences, as well as recruit other splicing factors.

1.3.3 Splice site selection

The consensus sequences of splice sites are short and relatively poorly conserved, especially the 3′ splice sites. The splice site selection becomes the key issue in precise splicing, especially from the alternative and/or suboptimal splice sites. Robberson *et al.* (1990) have proposed a model called "exon definition" in which exons are defined along the pre-mRNA before splicing (exon definition). After exons are defined, the adjacent exons are juxtaposed and spliceosome assembling is initiated. This mechanism will use the smallest distance between two adjacent splice sites as reference. In organisms with small introns, such as yeast and Drosophila, "intron definition" is proposed (Berget, 1995).

Splice site selection is also controlled by splicing enhancers. The interactions between SR proteins and splicing enhancers can promote the usage of alternative and/or suboptimal splice sites (Watakabe *et al.*, 1993). Different SR proteins interact with different enhancers in a sequence specific manner. The SR-enhancer complex may facilitate exon/intron definition and/or recruit the components of spliceosomes to the splice sites (Reed, 1996).

1.3.4 Alternative splicing

Alternative splicing of pre-mRNA is a common phenomenon in eukaryotic post-transcriptional modification. By alternative splicing, multiple mRNAs and, potentially, proteins with different functions can be generated from a single primary transcript in a tissue-specific manner. Regulation of splicing is very similar to the regulation of transcriptional initiation: *Trans*-acting regulatory proteins interact with *cis*-acting regulatory sequences of the pre-mRNA. By activating an alternative/suboptimal splice site, or by inhibiting a constitutively selected site, these interactions can modulate the splicing apparatus (the snRNP-pre-mRNA complex) and influence splice site selection (Adams *et al.*, 1996).

In vertebrates, chicken β-tropomyosin exons 6A and 6B are alternatively spliced in different tissues. Transcripts including 6A are present in non-muscle tissues and smooth muscles and transcripts including 6B are in skeletal muscles. Several regulatory elements and SR proteins are involved in the regulation of the alternative splicing. A G-rich enhancer downstream of 6B is necessary for efficient selection of 6B 5′ splice site. In addition, the ratio of different SR proteins in different tissues plays an important role in the alternative splicing (Gallego *et al.*, 1996, 1997; Pret and Fiszman, 1996).

A more complex example is expression of the transcriptional factor cAMP-responsive element modulator (CREM) gene which encodes a large family of important regulatory proteins (Lalli *et al.*, 1996; Lamas *et al.*, 1996). The CREM family includes three groups of isoforms, the cAMP-dependent transcriptional repressors (CREM α, β, γ) and activators (CREM τ, τ1, τ2, τα), and the inducible cAMP early repressor (ICER). The CREM activators are present in adult testis and the repressors are present in prepubertal testis. Both are transcribed from the non cAMP inducible GC-rich promoter P1 and differentiated by alternative splicing (Molina *et al.*, 1993; Stehle *et al.*, 1993). The ICER is transcribed from the cAMP inducible promoter P2 which is located in a 10kb intron in the middle of the gene. The ICER contains only CREM's basic DNA binding domain (DBD) and is a powerful repressor of many cAMP-induced transcriptions, including production of its own message (Molina *et al.*, 1993; Stehle *et al.*, 1993).

1.3.5 Introns and transcriptional initiation

As discussed in 1.2. gene expression is frequently governed by the interaction between multiple *cis*-acting DNA elements and *trans*-acting regulatory proteins. The *cis*-acting elements, especially the developmental and/or tissue-specific regulatory sequences, may also be present in introns. Those intronic regulatory elements include intronic promoters and transcriptional regulatory sequences (as well, of course, as splicing enhancers). These intronic elements are often involved in transcriptional initiation. In the extreme, an entire gene and, presumably, its regulatory sequences can reside within an intron. Two examples are the cuticle protein gene in Drosophila (Henikoff *et al.*, 1986a) and the RNA maturase genes in yeast (Kennel *et al.*, 1993; Schafer *et al.*, 1994; Grivell, 1996).

Intronic promoters are frequently tissue-specific promoters. Rat ca^{++}/calmodulin-dependent protein kinase IV (CaM kinase IV) is present mostly in thymic lymphocytes, cerebellum and testicular germ cells. CaM kinase IV gene contains 12 exons and 11 introns (Sun *et al.*, 1995). It encodes CaM kinase IV isoforms α and β and another protein calspermin, which is only present in postmeiotic male germ cells. The isoform β is transcribed from the 5′ end promoter and α is transcribed from the intronic promoter located in the first intron. The testis-specific calspermin promoter, as well as a testis-specific exon, is located in intron 10. This testis-specific promoter is exclusively expressed in postmeiotic testis of the transgenic mice (Sun and Means, 1995).

Another example is murine T-cell receptor (TCR) V α1 gene which contains a bi-directional intronic promoter. The transcript from the sense direction of the gene encodes a truncated V α1 gene product and the transcript from the antisense direction encodes an Ig β homologous protein (Maeda *et al.*, 1995).

Much of the best information on intronic transcriptional regulatory elements (apart from promoters) comes from *Drosophila melanogaster vestigial (vg)* gene, which encodes a protein required for normal wing formation. The expression of *vg* is initially in all cells of the embryonic wing disc. By second instar, this early universal expression ceases and *vg* expression is reactivated along the dorsoventral (D/V) boundary, then spreads to the whole disc again. This special expression is regulated by two separate intronic enhancers. The first enhancer is located in the second intron. It is 750bp long with a single transcription factor Suppressor of Hairless [Su(H)] binding site. Since it is required for dorsoventral (D/V) boundary formation in the second instar, it is named the D/V boundary enhancer. The [Su(H)] binding site is essential for the enhancer's function. The second enhancer is 806bp long and located in the fourth intron. It is necessary for late gene expression in the remainder of the wing and is called the quadrant enhancer. Reporter genes bearing the first enhancer element can only express in a restricted narrow stripe along the D/V boundary in the wing disc. The second enhancer cannot function without the normal D/V boundary expression. These two *vg* intronic enhancers play key roles in Drosophila wing development (Williams *et al.*, 1994; Kim *et al.*, 1996).

Some intronic regulatory elements function as silencers. In the rat osteocalcin (OC) gene, an intronic silencer is present in the first intron. This silencer suppresses the osteocalcin-CAT fusion gene expression approximately 10-fold. Mutation in the intron abolishes the effect. There is a protein binding motif TTTCTTT in this sequence, which may play the key role in the silencing event. The silencer retains partial effect when placed upstream of the OC promoter, so both the intronic environment and the sequence contribute to the silencing effect (Goto, *et al.*, 1996)

Intronic regulatory elements also work along with other regulatory elements. The *Drosophila melanogaster α1-tubulin* gene is expressed in almost all tissues. Three regulatory elements are responsible for this universal expression. Tubulin element 1 (TE1, 29bp) and tubulin element 2 (TE2, 68bp) are adjacent to the core promoter in the upstream regulatory region. The third element is located in the downstream intron (+144 to +635bp). Germ line transformation showed that these three elements successfully direct a *lac* Z reporter gene construct to express in all tissues. However, deletion or mutation of any one of these three elements results in line-to-line differences in tissue-specific expression pattern. This indicates that these DNA elements, including the intronic element, act collectively as general transcriptional activators in all tissues (O'Donnell *et al.*, 1994; O'Donnell and Wensink, 1994; Kohler *et al.*, 1996).

1.3.6 *Sex-lethal* in Drosophila

Regulation of expression of the *Sex-lethal (Sxl)* gene in *Drosophila melanogaster* is one of the best documented in eukaryotes. *Sxl* is the master regulatory gene which determines whether the fly follows the female or male developmental pathway by regulating the expression of genes involving sex determination and dosage compensation (Lucchesi and Skripsky, 1981; Cline 1984). It is presented here as a case study of how gene expression is regulated in a complex system.

Sxl is an RNA binding protein which regulates sex determination by promoting a cascade of sex-specific alternative splicing events leading to female sexual differentiation (Boggs *et al.*, 1987; Nagoshi *et al.*, 1988; Belote, 1989; Cline, 1993; Bopp *et al.*, 1996; Sakashita and Sakamoto, 1996). The *Sxl* gene is transcribed in both sexes but functional Sxl proteins are normally present only in female flies. In males, the *Sxl* primary transcripts are spliced through the default pattern resulting in an mRNA containing an extra exon, the male-specific exon. Since there is an in-frame stop codon in this male-specific exon, only short, non-functional polypeptides are synthesized in males. In females, the pre-existing Sxl protein binds to the pre-mRNA and the male-specific exon is skipped during splicing. As a result, the full-length functional Sxl protein can be produced in females (Sakamoto *et al.*, 1992; Horabin and Schedl, 1993). Thus Sxl proteins positively autoregulate their own expression by promoting the alternative splicing in female flies.

In general, synthesis of Sxl protein is dependent on the presence of functional Sxl protein. The question arises then, as to how the functional Sxl protein is present to initiate the autoregulatory loop in females? The *Sxl* gene contains two promoters, an early promoter and a late promoter (Salz *et al.*, 1989). The late promoter, also known as the maintenance promoter, is the main promoter used in both sexes. The early promoter, also known as female-specific promoter, is regulated by the X-chromosome-to-autosome (X:A) ratio and turned on transiently in the early developmental stage by 2X:2A (female) signal. The early transcript transcribed from the early-promoter is spliced bypassing the male-specific exon so functional Sxl proteins are synthesized (Zhu *et al.*, 1997). These early Sxl proteins direct the processing of *Sxl* pre-mRNA transcribed from the late-promoter into female mode. Once *Sxl* female splicing is established, it is maintained by the positive autoregulatory loop throughout life (Bell *et al.*, 1988; Cline, 1984; Keyes *et al.*, 1992). In males (1X:2A), the early-promoter is off so that the pre-mRNA transcribed from the late-promoter is spliced in the default male mode (Keyes *et al.*, 1992).

There is an elegant signalling system that registers the X:A ratio and regulates the early female-specific promoter. Genetic studies have revealed that three groups of genes are involved in the X:A signalling system. The first group is the numerator elements located on the X-chromosome, which includes *sisterless-a (sis-a)*, *sisterless-b (sis-b)*, *sisterless-c (sis-c)* and *runt (run)*. The expression of numerator genes is required for the zygote to count X-chromosomes. Loss of function mutations of these genes affect the selection of female pathway in early developmental stage in 2X:2A females but have no effect in the 1X:2A males (Duffy and Gergen, 1991; Torres and Sánchez, 1992; Cline, 1993). The second group is denominator elements on autosomes. In contrast to the numerator function, loss of denominator function results in failure to choose the male pathway in 1X:2A animals. Only one denominator gene, *deadpan (dpn)*, on the second chromosome has been identified so far (Younger-Shepherd *et al.*, 1992). The third group consists of maternal genes *daughterless (da)* and *extramachrochaetae (emc)*. *da* acts as a cofactor with the numerator proteins and *emc* interacts with the denominator proteins (Cline, 1993; Barbash and Cline, 1995).

Molecular studies show that these X:A signalling genes encode transcriptional regulatory proteins which regulate transcription from the *Sxl* early promoter (Estes *et al.*, 1995; Hoshijima *et al.*, 1995). There is a 1.4kb fragment upstream from the early-promoter containing all the *cis*-regulatory DNA elements responsible for the X:A signal (Estes *et al.*, 1995). The regulatory protein Da/Sis-b heterodimer activates the early-promoter by directly binding to this regulatory sequence. In contrast, Dpn protein acts as the repressor and represses the Da/Sis-b dependent activation by binding to a different binding site. Another repressor, Emc, works in a different manner. It inhibits Da/Sis-b heterodimer formation, probably by competing with either Da or Sis-b proteins. Thus, the regulation of early Sxl expression by the X:A signal system seems to be achieved by the balance between numerators and denominators (Hoshijima *et al.*, 1995; Figure 1-1).

In addition to its own expression, Sxl also regulates the pre-mRNA processing of next gene in the sex-determination cascade, *transformer (tra)*. In males, *tra* mRNA contains a premature stop codon so no functional Tra protein is produced. In females, Sxl proteins bind to the polypyrimidine tracts on the upstream of the regulated 3′ splice site (upstream of the stop codon) and block splicing proteins access to

Page 21 has been removed due to copyright restrictions. The information removed was Figure 1-1 Model for the regulation of *Sxl* geneModel for the regulation of *Sxl* gene adapted from Estes *et al.*, 1995.

this 3′ site. This results in the usage of the weaker female-specific 3′ splice site downstream (also downstream from the stop codon) and skips the otherwise in-frame stop codon. Thus, functional Tra protein can be synthesized from female mRNA (Adams *et al.*, 1996; Granadino *et al.*, 1997). In turn, Tra protein, associated with other splicing proteins, modulates the alternative splicing of pre-mRNA of *doublesex (dsx)*, another gene in sex determination cascade, by binding with its exonic splicing enhancer and activating the suboptimal female specific 3' splice sites (Tian and Maniatis, 1993; 1994; Lynch and Maniatis, 1996).

Sxl also regulates dosage compensation, which equalizes the production of X-linked gene between males and females, through a different mechanism. Four autosomal genes, *maleless (mle)* and *male-specific lethal 1, 2* and *3 (msl-1, msl-2 and msl-3)*, are known to be required for dosage compensation (Baker *et al.*, 1994; Zhou *et al.*, 1995). Loss of function of these genes results in male lethality because of the failure to achieve dosage compensation, but has no effect in females. The MLE and MSL proteins act as a multi-subunit complex which is preferentially associated with numerous sites along the X chromosome in males. All four wild-type proteins must be simultaneously present for the normal function of the MLE/MSL complex (Palmer *et al.*, 1993; Baker *et al.*, 1994; Lucchesi, 1996). The binding of MLE/MSL complex to the X chromosome is correlated to the increases of histone acetylation which may remodel the chromatin structure and result in a doubling of transcription.

Sxl protein regulates dosage compensation through regulating the translation of the key Msl protein, Msl-2. With the RNA binding function, Sxl binds to multiple sites along the 5′ and 3′ untranslated region of *msl-2* transcripts. This binding represses the translation of *msl-2* mRNA. Thus, Msl-2 protein will not be produced or produced at a low level in females so that no dosage compensation will occur. Sxl may also partially repress the expression of the *msl-1* gene through the same mechanism in females. Since there is no Sxl protein in males, Msl-2 and Msl-1 proteins will be produced and dosage compensation does occur (Bashaw and Baker, 1997; Kelley *et al.*, 1997).

As mentioned above, alternative transcriptional initiation, alternative splicing and, probably, translational regulation, are essential to the proper function of the regulation of the sexual developmental pathway. It has, as well, ramifications for our understanding of the role of chromatin structure in gene regulation. Thus, this single example encompasses most of the known means of eukaryotic gene regulation and demonstrates how multiple regulatory mechanisms can be harmoniously and precisely employed in a given system to fulfill its different biological functions.

1.4 Purine metabolism as a model system for studying gene regulation

1.4.1 An introduction to purine metabolism

1.4.1.1 Purines

Purines are compounds containing 5 and 6 membered heterocyclic rings and are predominantly present as nucleotides and polynucleotides (DNA and RNA) in the cell. They are involved in many cell functions, such as energy metabolism, signal transduction and cofactors in enzyme reactions (Henderson, 1972; Patterson, 1985). In energy metabolism, purine nucleotides act as the energy carriers. The energy released from the hydrolysis of ATP and GTP, generally to ADP and GDP, is used to drive energy consuming reactions. In signal transduction, they act as the second messengers in the form of cyclic nucleotides. In enzyme reactions, purines work as co-factors of the group-transfer co-enzymes which transfer compounds such as sugars, sulfates, alkyl groups, acyl groups and hydrides. Purines are also directly involved as reactants in the biosynthesis of histidine, riboflavin, folic acid, thiamine, glycolipids, glycoproteins and pteridines. The latter are a major component of pigments in many animals, and, of particular importance to this thesis, specifically in Drosophila eyes (Henderson, 1972; Nash and Henderson, 1982; Patterson, 1985). Perhaps most importantly, they are essential components of DNA and RNA, which carry genetic information.

1.4.1.2 *de novo* purine nucleotide biosynthesis

Most organisms can synthesize purine nucleotides *de novo* from simpler carbon and nitrogen compounds (Figure 1-2; Table 1-2). It takes ten steps to synthesize inosine monophosphate (IMP), the first complete purine nucleotide and the major branch point of the pathway (Flaks and Lukens, 1963; Henderson, 1972). IMP is the last common precursor of adenosine monophosphate (AMP) and guanosine monophosphate (GMP). Two, two-step pathways lead to the formation of AMP and GMP from IMP (Henderson *et al.*, 1977). Subsequently, the nucleoside monophosphates can be phosphorylated to the respective di- and triphosphylated forms. Reduction of ribonucleoside diphosphates by ribonucleotide reductase generates deoxyribonucleoside diphosphates, which are needed primarily for DNA synthesis (Figure 1-3).

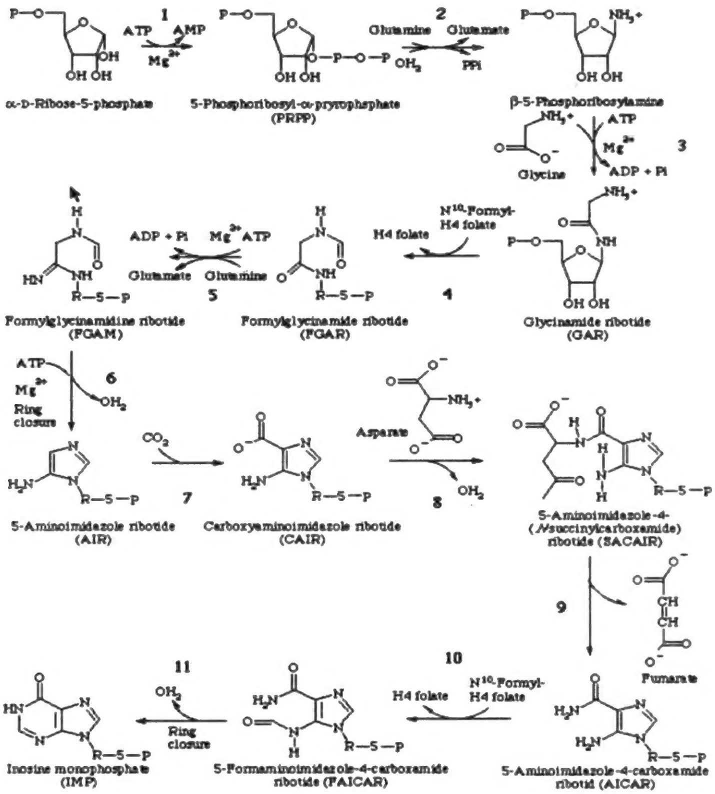
1.4.1.3 Purine salvage pathways

Purine nucleotides can also be generated by a series of "salvage reactions" which require much less energy than the *de novo* pathway. Purine bases derived from nutrition or from breakdown of cellular nucleic acids (particularly from RNA) can be reconverted to their corresponding nucleotides by phosphoribosylation (Figure 1-3). Adenine can be converted to AMP by adenine phosphoribosyltransferase (APRT), guanine and hypoxanthine can be converted to GMP and IMP, respectively, by hypoxanthine-guanine phosphoribosyltransferase (HGPRT). The purine nucleosides adenosine, guanosine and inosine can be phosphorylated to their respective nucleotides by kinases (Henderson, 1972; Nash and Henderson, 1982).

1.4.1.4 Purine interconversions

Purine nucleotides (AMP, GMP and IMP) can be interconverted in many organisms (Henderson *et al.*, 1978). AMP and, sometimes, GMP can be reconverted to their common precursor IMP, by adenylate deaminase and guanylate reductase respectively (Figure 1-3). However, guanylate reductase seems not to be present in some organisms, of which both yeast and *Drosophila melanogaster* are examples (Nash and Henderson, 1982).

Figure 1-2 The pathway of *de novo* IMP biosynthesis



The enzyme catalyzing each of the steps are as follows:

1. ribose phosphate pyrophosphokinase
2. amidophosphoribosyl transferase
3. phosphoribosylamine--glycine ligase (GAR synthetase, GARS)
4. phosphoribosylglycinamide formyltransferase (GAR transformylase, GART)
5. phosphoribosylformylglycinamidine synthase (FGAM synthetase, FGARAT)
6. phosphoribosylformylglycinamidine cyclo-ligase (AIR synthetase)
7. phosphoribosylaminoimidazole carboxylase (AIR carboxylase)
8. phosphoribosylaminoimidazole-succinocarboxamide synthase (SACAIR synthetase)
9. adenylosuccinate lyase
10. phosphoribosylaminoimidazole carboxamide formyltransferase (AICAR transformaylase)
11. IMP cyclohydrolase (IMP synthetase)

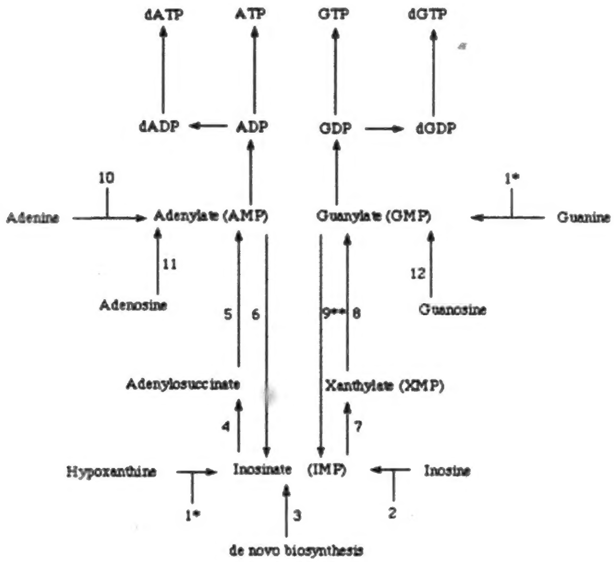
(Adapted from Voet and Voet, 1995)

Table 1-2 Purine biosynthesis involved enzymes and genes

		Name of the genes				
Name of the enzymes	Classification number	*E. coli*	*B. subtilis*	*S. cerevisiae*	*D. melanogaster*	*H. sapiens*
amidophosphoribosyl transferase	EC: 2.4.2.14	purF	purF	ADE4	Prat	GPAT
GARS	EC: 6.3.4.13	PurD	purD	ADE5, 7	ade3 (GART)	GART
GART	EC: 2.1.2.2	purN	purN	ADE8	ade3 (GART)	GART
FGAM synthetase	EC: 6.3.5.3	purL	purLQ	ADE6	ade2	FGAMS
AIR synthetase	EC: 6.3.3.1	purM	purM	ADE5, 7	ade3 (GART)	GART
AIR carboxylase	EC: 4.1.1.21	purEK	purK	ADE2		AIRC
SACAIR synthetase	EC: 6.3.2.6	purC	purC	ADE1		AIRC
adenylosuccinate lyase	EC: 4.3.2.2	purB	purB	ADE13		ASL
AICAR transformylase	EC: 2.1.2.3	purH	purH	ADE16/ADE17		IMPS
IMP synthetase	EC: 3.5.4.10	purH	purH	ADE16/ADE17		IMPS
adenylosuccinate synthetase	EC: 6.3.4.4	purA	purA	ADE12		ADSS
adenylosuccinate lyase	EC: 4.3.2.2	purB	purB	ADE13		ASL
IMP dehydrogenase	EC: 1.1.1.205	guaB	guaB	PUR5	ras	IMPDH1 and 2
GMP synthase	EC: 6.3.4.1	guaA	guaA	GUA1	bur	GMPS

* EC: Enzyme Commission

Figure 1-3 Purine salvage pathway and interconversions



Salvage pathways and interconversions of purine nucleotides. The names of the enzymes catalyzing each of the numbered steps are as follows:

1. hypoxanthine-guanine phosphoribosyltransferase (HGPRT)*
2. inosine kinase
3. de novo biosynthetic pathway
4. adenylosuccinate synthetase
5. adenylosuccinate lyase
6. adenylate deaminase
7. inosinate dehydrogenase
8. guanylate synthetase (xanthylate aminase)
9. guanylate reductase**
10. adenine phosphoribaosy transferase (APRT)
11. adenosine kinase
12. guanosine kinase

-nucleotide monophosphates are phosphorylated to form di- and triphosphates
-ribonucleotide diphosphates can be reduced to form deoxyribonucleotides

* this pathway has weak activity in Drosophila
** this pathway is probably absent from Drosophila

(Adapted from Leonard, 1986).

1.4.1.5 Purine catabolism

The cellular purines and purine derivatives are converted to excretory end products via four processes, dephosphorylation, deamination, phosphorolysis and oxidation (Figure 1-4).

1.4.1.6 Rate limiting steps in purine biosynthesis

Purine biosynthesis is under tight regulatory controls (Henderson, 1972). The most important regulatory step in *de novo* pathway is the first reaction, phosphoribosyl pyrophosphate (PRPP) to phosphoribosylamine, catalyzed by glutamine PRPP amidotransferase (Smith *et al.*, 1994; Zhou *et al.*, 1994). This step represents commitment to purine biosynthesis and is the rate-limiting step of the *de novo* pathway. Glutamine PRPP amidotransferase is feedback-inhibited by the final products, the purine nucleotides, especially AMP and GMP (Smith *et al.*, 1994; Zhou *et al.*, 1994). In most organisms, the enzyme activity is effectively inhibited by AMP, GMP and IMP. On the other hand, the substrates glutamine and PRPP can activate the reaction. The rate of the reaction is determined by the balance between the substrates and final products (Bashkin and Sperling, 1978; Holland, *et al.*, 1978; Taylor *et al.*, 1982).

Another regulatory point is in the branch pathway from IMP to AMP and IMP to GMP. Adenylosuccinate synthetase catalyzes the first step from IMP to AMP and is feedback-inhibited by AMP. IMP dehydrogenase catalyzes the first step from IMP to GMP and is feedback-inhibited by GMP. Both enzymes are rate-limiting to the pathway in which they operate. Since ATP is required for GMP synthesis and GTP is required for AMP synthesis, increased ATP level accelerates GMP synthesis, and *vice versa*. In some cases, excess AMP may be converted to IMP, which serves as the source of GMP synthesis.

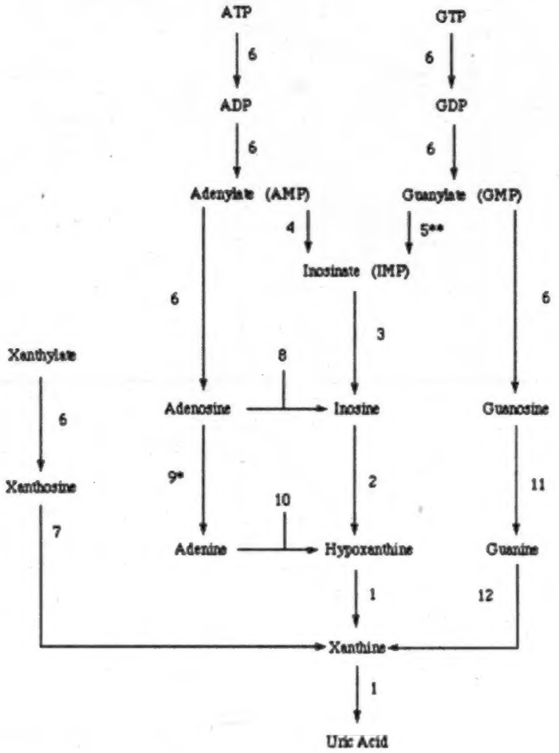
The above three enzymes are the key enzymes in the regulation of the purine biosynthesis pathway, which can be achieved by activating or inhibiting the activities of these enzymes, and/or increasing enzyme concentration via transcriptional or translational regulation (Patterson, 1985).

1.4.2 Purine biosynthesis in bacteria

1.4.2.1 *Escherichia coli*

In *E. coli*, six monocistronic and four small polycistronic operons encode the enzymes utilized in purine nucleotide biosynthesis (Zalkin and Dixon, 1992). Seven of them, purF, purL, purMN, purEK, purHD, purC and purB, encode enzymes for *de novo* IMP biosynthesis; two of them, purA and purB, encode enzymes for the conversion of IMP to AMP and the operon guaBA encodes enzymes for the conversion of IMP to GMP (Table 1-2). Notice that the single enzyme encoded by purB, adenylosuccinate lyase, is also involved in *de novo* IMP biosynthesis. The last operon, purR, encodes the purine repressor protein PurR, a transcriptional regulator, which has an evolutionary relationship to the lac repressor protein (Schumacher *et al.*, 1994a, b; Sauer, 1996).

Figure 1-4 Purine catabolism



The names of the enzymes and/or the type of reaction involved at each step are as follows:

1. xanthine dehydrogenase (XDH) - oxidation reaction
2. inosine phosphorylase - phosphorolysis reaction
3. dephosphorylation reaction
4. adenylate deaminase - deamination reaction
5. guanylate deaminase (reductase)** - deamination reaction
6. dephosphorylation reactions
7. xanthine phosphorylase - phosphorolysis reaction
8. adenosine deaminase - deamination reaction
9. adenosine phosphorylase - phosphorolysis reaction*
10. adenine deaminase - deamination reaction
11. guanosine phosphorylase - phosphorolysis reaction
12. guanine deaminase - deamination reaction

* this pathway has weak activity in Drosophila
** this pathway is probably absent from Drosophila

(Adapted from Leonard, 1986).

All the purine operons, and some other genes related to nucleotide metabolism, are co-regulated by PurR. They form the *pur* "regulon" and share a conserved 16bp AT rich pseudo-palindromic operator which is the binding site for PurR protein. The binding of PurR to the operators requires the presence of a co-repressor, either hypoxanthine or guanine (Rolfes and Zalkin, 1990). The interaction between *pur* operators and PurR is the key regulatory mechanism of the *pur* regulon, because mutations of a *pur* operator or PurR can result in high level constitutive expression of the corresponding operon or, in the latter case, of the whole regulon.

PurR is transcriptionally auto-regulated (Meng *et al.*, 1990; Rolfes and Zalkin, 1990; Zalkin and Dixon, 1992). There are two operators in the PurR gene. Operator O_1 is located between the transcription start site and the translation start site, while operator O_2 is in the coding region. A series of plasmids containing the purR-lacZ reporter gene was constructed and transformed into both $PurR^+$ and $PurR^-$ cells. It was found that in the $PurR^+$ cells, the expression of the lacZ reporter gene was repressed two to three times when purine (either adenine or hypoxanthine) was present. This repression was not observed in the $PurR^-$ cells (Meng *et al.*, 1990; Rolfes and Zalkin, 1990). Both O_1 and O_2 are required for this autoregulation although no evidence of cooperation between them was found. The binding between PurR and these operators is dependent on the concentration of PurR (Rolfes and Zalkin, 1990).

PurR is a homodimer with an N-terminal DNA-binding domain and a C-terminal co-repressor-binding and dimerization domain (Rolfes and Zalkin, 1988; Schumacher *et al.*, 1994a, b). The DNA-binding domain of each monomer can bind one half-site of the *pur* operator. Unlike LacR, PurR cannot bind the operator by itself. The binding of PurR and the co-repressor changes the structure of PurR and allows it to bind the operator. When co-repressor is not available, PurR dissociates from the operator and transcription is initiated. Thus the expression of genes related to purine biosynthesis is regulated by the concentration of hypoxanthine and guanine, which must act as indicators of the availability of purine sources (Schumacher *et al.*, 1995; Sauer, 1996).

1.4.2.2 *Bacillus subtilis*

In *B. subtilis*, the genes that encode the enzymes for purine biosynthesis are organized differently from those in *E. coli*. The 11 genes involved in *de novo* biosynthesis of IMP are clustered within a 12-cistron *pur* operon (Ebbole and Zalkin, 1987). The gene order of *B. subtilis pur* operon is purEKB-C(orf)QLF-MNH-D, in which "-" indicates the "intercistronic regions" discussed further below, and "orf" refers to an open reading frame with unknown function. The combined function of purQ and purL is equivalent to the single *E. coli* purL polypeptide, which encodes FGAM synthetase. The *pur* operon is transcribed from upstream of purE to downstream of purD.

The *pur* mRNA is separated into three translationally "overlapping groups" by the "intercistronic regions", which are regions where the DNA sequence does not encode an amino acid sequence. In contrast, within each overlapping group, the 3′ end of the upstream coding sequence overlaps the 5′ end of the next coding sequence (Ebbole and Zalkin, 1989). Gene overlapping seems to be a common feature of the biosynthetic pathway for related operons in *B. subtilis* and other species of Bacillus (Henner *et al.*, 1985; Quinn *et al.*, 1991; Zalkin and Dixon, 1992; Christiansen *et al.*, 1997). It is thought that gene overlap generates translational coupling which provides the mechanism for synthesis of the subunits of a functional complex at 1:1 ratio from a polycistronic mRNA (Oppenheim and Yanofsky, 1980; Zalkin and Ebbole, 1988).

Transcription of the *B. subtilis pur* operon is regulated by the concentration of adenine and guanine through two independent mechanisms acting, respectively, at transcriptional initiation and during transcriptional elongation/termination (Ebbole and Zalkin, 1988; Zalkin and Ebbole, 1988; Zalkin and Dixon, 1992). Adding adenine to the medium prevents transcriptional initiation and adding guanine to the medium results in an mRNA prematurely terminated just before purE. The first mechanism is not as clear as the equivalent control mechanism in *E. coli*. *B. subtilis* repressor PurR has been cloned and sequenced (Weng *et al.*, 1995). Experiments show that a high concentration of PRPP can cause PurR to disassociate from the *pur* operator and facilitate transcription initiation. The second mechanism is well studied (Ebbole and Zalkin, 1988). There are two overlapping dyad symmetries in the initial ~200bp "leader sequence" of *pur* mRNA. These two dyads have the potential to form either of two possible RNA base pairing motifs (hairpins), an anti-terminator or a terminator. When guanine is present, it activates a guanine responsive regulatory protein, which binds to the DNA in the anti-terminator forming region. This allows the terminator base-paired secondary structure to be formed and transcription is terminated. When there is insufficient guanine available to activate the regulatory protein, it is inactive and cannot bind the antiterminator, thereby favoring anti-terminator formation and promoting full-length transcription. Antitermination/termination regulation of this kind is known as attenuation (Shimotsu *et al.*, 1986).

In both *E. coli* and *B. subtilis*, all the genes involved in purine biosynthesis have been cloned and sequenced. As mentioned above, the concentration of purines not only regulates the enzyme activities, but also the expression of the related genes.

1.4.3 Purine biosynthesis in eukaryotes

1.4.3.1 *Saccharomyces cerevisiae*

In *Saccharomyces cerevisiae* (Baker's yeast), all the genes encoding enzymes related to purine nucleotide biosynthesis have been cloned (Tibbetts and Appling, 1997). Among these genes, *ADE5,7* encodes the bifunctional enzyme GARS/AIRS which catalyzes the second and fifth steps (Henikoff, 1986). Two genes, *ADE16* and *ADE17*, encode isoforms of another bifunctional enzyme, AICART/IMPCH, which catalyzes the ninth and tenth steps (Tibbetts and Appling, 1997). The amino acid sequences of these two isozymes are highly conserved and the functional relationship between them remains under investigation. Four individual genes located on at least two different chromosomes have been reported to encode isoforms of IMP dehydrogenase (Johnston *et al.*, 1994; Bussey *et al.*, 1995). It has been observed that the transcription of *ADE5,7*, *ADE8*, *ADE4*, *ADE1* and *ADE2* is increased during adenine starvation and decreased when the exogenous adenine is available (Daignan-Fornier and Fink, 1992; Rolfes and Hinnebusch, 1993; Stotz *et al.*, 1993; Rolfes *et al.*, 1997). Thus it seems that the purine biosynthetic genes are co-ordinately regulated by adenine at the transcriptional level. The adenine-mediated regulation is related to the transcription regulating proteins BAS1 and BAS2 (BAS refers to "basal level control", Arndt *et al.*, 1987) because both are required for transcription of the genes in adenine-free medium and the transcription of *ADE* genes is low and unregulated in a *BAS1* and/or *BAS2* mutant background (Daignan-Fornier and Fink, 1992; Rolfes and Hinnebusch, 1993; Gedvilaite and Sasnauskas, 1994; Stotz *et al.*, 1993).

BAS1 and BAS2 are DNA binding proteins that are known to bind the Ade upstream activation sequences (UAS_{Ade}) of *ADE2* and *ADE5,7* genes under appropriate conditions. Thus, when adenine is not available, BAS1 and BAS2 bind to each other and form a protein complex which binds the UAS_{Ade} (Zhang *et al.*, 1997). This binding activates the transcription of both *ADE2* and *ADE5,7* (Daignan-Fornier and Fink, 1992; Rolfes *et al.*, 1997). When adenine is available, it associates with BAS1, changing its conformation and masking the protein binding and activating domains of BAS1. As a result the BAS1/BAS2 complex disassociates or is not formed, leaving the *ADE* genes inactive (Zhang *et al.*, 1997).

1.4.3.2 *Drosophila melanogaster*

In *Drosophila melanogaster*, five genes encoding seven enzymes of purine nucleotide biosynthesis have been cloned so far. Among them, *ade3* (GART) encodes a trifunctional enzyme GARS/GART/AIRS (Henikoff *et al.*, 1986b); *ade2* encodes FGARAT (Tiong and Nash, 1993); *Prat* encodes PRAT (Clark, 1994); *ras* encodes IMPDH (Nash *et al.*, 1994) and *bur* encodes GMPS (Chomey and Nash, 1995). Recently, the gene encoding the bifunctional enzyme AIRC-SAICARS was cloned and located at X11B4-14, approximately the same location as *ade4* and *ade5* mutations (O'Donnell and Clark, 1997).

The structure of the GART *(ade3)* gene is complex. It is about 10kb and encodes a 9kb transcript. There are six introns in the primary transcript. An additional mRNA 3′ processing site is present in the fourth intron. Thus, by using the alternative mRNA termination sites, two transcripts can be produced from the *ade3* gene (Henikoff *et al.*, 1983; Henikoff *et al.*, 1986b). One encodes the first enzyme, GARS, the other encodes the trifunctional enzyme. Most importantly, a cuticle gene is nested in the first intron and transcribed from the opposite direction (Henikoff *et al.*, 1986a). There is no functional relationship between the GARS/GART/AIRS enzyme and the cuticle protein. The structure of *ade3* (including the nested gene) is conserved in *Drosophila pseudoobscura,* a species relatively distantly related to *Drosophila melanogaster* (Figure 1-5, Henikoff and Eghtedarzadeh, 1987). However, neither the nested cuticle gene nor the alternative mRNA termination site are present in the equivalent DNA sequence in *Chironomus tentans* (Clark and Henikoff, 1992).

The trifunctional enzyme encoded at the *ade3* locus is also present in vertebrates (Zalkin, 1993). Multifunctional polypeptides involved in catalyzing several related biochemical steps are often found in biosynthetic pathways. In addition, functionally related enzymes can often be co-purified as an enzyme complex, even in situations where they are not vested in a single polypeptide. One advantage of multifunctional enzyme complexes is believed to be that they provide co-ordination for linked enzymatic functions (Henikoff, 1987; Zalkin, 1993). Zalkin (1993) postulated that a complex of PRAT, GART and FGARAT could couple the first five steps of *de novo* purine biosynthesis and produce the fifth product from PRPP without releasing any intermediates.

Page 32 has been removed due to copyright restrictions. The information removed was Figure 1-5 Structure of Drosophila *ade3* (GART) gene adapted from Henikoff and Eghtedarzadeh, 1987.

The FGARAT gene is located in a 7.1kb *Eco* RI genomic fragment of the *ade2* locus. When using the 7.1kb *Eco* RI fragment as a probe, two transcripts, 2.7kb and 4.8kb, were found in Northern analyses of the wild-type poly-A^+ RNA. Eight cDNA clones were isolated. A comparison between the longest cDNA (3.7kb) and the genomic sequences revealed that there are two introns (75bp and 62bp) in the transcribed region (Tiong and Nash, 1993).

The GMPS gene is located in the *bur* locus on the second chromosome (Chomey and Nash, 1995). The gene encodes an approximate 9.5kb primary transcript. A 3.4kb cDNA was isolated. Sequence comparison revealed that there are at least seven introns in the primary transcript, ranging from 50bp to 1.8kb.

The *Prat* gene is located in a 3.5kb *Eco* RI genomic fragment. Unlike the above genes, there is no intron interrupting the coding region in the 2.1kb *Prat* transcript (Clark, 1994). Purine metabolism in *Drosophila melanogaster* will be discussed further in chapter 1.5.

1.4.3.3 *Homo sapiens*

cDNAs which encode all of the enzymes involved in human purine biosynthesis have been isolated (Zalkin and Dixon, 1992; Zalkin, 1993). Three of them encode multifunctional enzymes. *GART* encodes a trifunctional enzyme GARS/GART/AIRS, *AIRC* encodes a bifunctional enzyme AIRC/SAICARS and *IMPS* encodes yet another bifunctional enzyme AICART/IMPCH (Zalkin, 1993). Two cDNAs encode the functionally indistinguishable but structurally distinct isoforms of IMPDH, IMPDH type I and IMPDH type II (Collart and Huberman, 1988; Natsumeda *et al.*, 1990). Recently, four genes (*GPAT* or *PRAT*, *AIRC* and the two IMPDH genes) have been cloned (Zimmermann *et al.*, 1995, 1996; Chen *et al.*, 1997).

The activity of IMPDH and the expression of IMPDH genes are strongly linked to cellular proliferation, differentiation and the availability of purine nucleotides (Zimmermann *et al.*, 1996). The IMPDH I gene is transcribed constantly at a low level whereas the transcription of the IMPDH II gene is at high levels mainly in dividing (including neoplastic), cells (Konno *et al.*, 1991; Nagai *et al.*, 1991, 1992). Senda and Natsumeda (1994) have reported that both isozymes are expressed in a wide range of tissues but at different levels.

The IMPDH I gene contains three different promoters which are used in different cell types and different cell developmental stages (Gu *et al.*, 1997). The IMPDH II gene contains a 466bp 5′ regulatory sequence including the transcription and translation start sites. Several transcription-regulating protein binding sites have been identified within this sequence. This 5′ end sequence is able to initiate transcription of a CAT reporter gene and has showed an increased expression when the host cells enter cellular proliferation (Zimmermann *et al.*, 1996).

GPAT and *AIRC* genes are adjacent to each other. They share a 558bp intergenic promoter region and are transcribed divergently (Chen *et al.*, 1997). No TATA or CCAAT motif has been found in the promoter region. Bi-directional promoters are also known to be associated with several other pairs of functionally related genes (Lennard and Fried, 1991; Fischer *et al.*, 1993; Ohshige *et al.*, 1993; Zhu *et al.*, 1994).

The activities of purine biosynthetic enzymes and the expression of the related genes are clearly related to the availability of purine pools in eukaryotes. However, little is known about the regulation of these genes at the molecular level.

1.5 Drosophila purine genetics

1.5.1 Purine metabolism in Drosophila

1.5.1.1 Purine biosynthesis

Purine metabolism in *Drosophila melanogaster* has been studied intensively (Henderson and Scott, 1980; Henderson *et al.*, 1980; Nash and Henderson, 1982). Purines are not required in the fruit-fly's diet (Sang, 1956; Geer 1964). Cell cultures show that *Drosophila melanogaster* cells can grow in purine-free media and adding inhibitors of *de novo* purine biosynthesis blocks growth (Becker, 1974 a, b; Wyss, 1977). Tracer studies show that radioactive IMP, AMP and GMP are found in larvae growing in media containing labeled purine precursors (Johnson, 1978; Johnson *et al.*, 1980 a, b, c). This evidence demonstrates that a *bona fide* purine nucleotide biosynthetic pathway is present in *Drosophila melanogaster* (Nash and Henderson, 1982; Henikoff *et al.*, 1986c; Tiong and Nash, 1993; Clark, 1994).

It has been observed that dietary purine bases can stimulate the growth of wild-type flies (Villee and Bissell, 1948; Hinton *et al.*, 1951; House, 1962; Sang, 1978) and rescue purine auxotrophic mutants (Nash and Henderson, 1982) while an excess of exogenous purines is toxic to flies (Henderson and Scott, 1980; Henderson *et al.*, 1980). These observations indicate that free purine bases can be converted to purine nucleotides, so purine salvage pathways must exist in Drosophila (Nash and Henderson, 1982). Indeed, high levels of APRT activity are found in flies (McMaster-Kaye and Taylor, 1959; Wyss, 1977). However, the activity of the other salvage enzyme, HGPRT, is absent or only at a very low level in Drosophila cell cultures (Becker, 1974 a) and larvae (Johnson *et al.*, 1980 b, c). In addition, hypoxanthine cannot rescue cell cultures treated with inhibitors of *de novo* purine biosynthesis, but the same cells can grow when adenine and adenosine are added (Wyss, 1977). It has been suggested that the lack (or low level) of the HGPRT reaction product may be the result of the high activities of xanthine dehydrogenase and guanine deaminase which convert hypoxanthine and guanine to xanthine (Nash and Henderson, 1982), although no experimental evidence has been produced to corroborate this suggestion. In contrast, nucleoside kinases seem to be active in Drosophila because purine nucleosides can be converted to purine nucleotides (Becker, 1974 a, b; 1975; Wyss, 1977; Johnson *et al.*, 1980 a, b, c).

In *Drosophila melanogaster*, AMP can be converted to other purine nucleotides but GMP is barely converted (Nash and Henderson, 1982). Of the two interconversion related enzymes, adenylate deaminase activity was detected, but little activity of guanylate reductase was found in embryonic cell lines of *Drosophila melanogaster* (Becker, 1974 b). Tracer experiments showed that after larvae were fed radioactive adenine, labeled AMP, IMP, XMP and GMP were found in the larval extracts (Hodge and Glassman, 1967; Johnson *et al.*, 1980 b, c). The radioactive IMP is presumably derived from AMP synthesized from the labeled adenine through a salvage pathway. However, when guanine was used as precursor, little labeled AMP, IMP and XMP were found (Johnson *et al.*, 1980 b, c). Since the rate of conversion of GMP to other purine nucleotides is limited by the low activity of guanylate reductase, guanine and guanosine cannot be used to rescue the mutants with blocked *de novo* purine biosynthesis. In contrast, adenine, adenosine and inosine are usable supplements to the diet of such mutants, which is useful in identification of mutants defective in purine biosynthesis (Nash and Henderson, 1982).

1.5.1.2 Pteridine metabolism in Drosophila

Pteridines are purine derivatives and carry out a variety of functions in most organisms (Werner *et al.*, 1996). In Drosophila, pteridines form eye pigments which affect eye-colour. Drosophila eye-colour is produced by the combination of the brown ommochromes and pteridines. Most of the wild-type pteridine eye pigments are the orange-red drosopterins, with a small proportion of the yellow sepiapterin (Ziegler, 1961; Ziegler and Harmsen, 1969).

Pteridines are compounds derived from GTP (Brown, 1971). They share a common pyrimidine-pyrazine ring. The basic structure of pteridines is illustrated by 2-amino-4-hydroxypteridine, also known as "pterin" (Figure 1-6a; Ziegler and Harmsen, 1969). Drosopterins are dimeric pterins (Rokos and Pfleiderer, 1975). The pteridine biosynthetic pathway is not yet fully understood. The first and the rate-limiting step of pteridine biosynthesis is the conversion of GTP to dihydroneopterin triphosphate by GTP cyclohydrolase (Burg and Brown, 1966; 1968; Fan *et al.*, 1976). In *Drosophila melanogaster*, GTP cyclohydrolase is encoded at the *Punch* locus (MacKay and O'Donnell, 1983). Dihydroneopterin triphosphate is quickly converted to sepiapterin and other pteridines (Figure 1-6b, Fan *et al.*, 1975; Yim *et al.*, 1977; Fan and Brown, 1979; Krivi and Brown, 1979; Dorsett and Jacobson, 1982).

Pteridines are particularly of interest in this thesis because interruptions of purine biosynthesis (probably GMP synthesis) result in anomalous eye-colour. There are at least 30 different Drosophila eye-colour mutants which have reduced or absent levels of pteridines (Hadorn and Mitchell, 1951; Nolte, 1954 a, b; 1955, 1959; Forrest *et al.*, 1956; Counce, 1957; Gregg and Smuckler, 1965; Baker, 1973). Many of these mutants have been studied to determine the gene-enzyme relationship in pteridine metabolism (Fan *et al.*, 1976; Krivi and Brown, 1979). As we will see, the study of one of them, the purplish eye-colour mutants of the *raspberry (ras)* locus which has been found to produce altered pteridine metabolism and GTP cyclohydrolase activities (Fan *et al.*, 1976; Evans and Howells, 1978; Fan and Brown, 1976, 1979) led to the identification of the gene which encodes the enzyme IMP dehydrogenase, the first and rate limiting enzyme in GMP synthesis (Nash *et al.*, 1994).

Page 36 has been removed due to copyright restrictions. The information removed was Figure 1-6 Structure of pterin adapted from Nash and Henderson, 1982; Dorsett and Jacobson, 1982.

1.5.2 Auxotrophs in purine biosynthesis

1.5.2.1 Auxotrophs in *Drosophila melanogaster*

Auxotrophs are mutants requiring exogenous compounds which are normally synthesized endogenously in non-mutant individuals. The mutations in auxotrophs are often related to genes encoding the enzymes of biosynthetic pathways, resulting in blocked normal biosynthesis. If a mutant has reduced enzyme activities, rather than none, it is generally called a "leaky" mutant. A "leaky" mutant may well show delayed development and/or only relatively low viability, rather than classical auxotrophy, in the absence of nutritional supplementation. In principle, however, a seriously defective mutant, that is leaky at the enzyme level, can behave as an auxotroph (Nash and Henderson, 1982).

The first induced auxotrophic mutants in *Drosophila melanogaster* were recovered by Vyse and Nash (1969) and one of them turned out to be an adenosine and pyrimidine double requirement (Vyse and Sang, 1971). Since then, a number of purine auxotrophs have been isolated and used to identify and clone the relevant genes. According to the dietary purine supplements, these mutants can be defined as adenosine requiring, guanosine requiring or purine requiring auxotrophs. As explained above, as a result of the lack of GMP reductase activity in *Drosophila melanogaster*, guanine sources can only be used to rescue mutants with a block in production of IMP from GMP, whereas, adenine sources are appropriate supplements for mutations in *de novo* IMP biosynthesis and conversion from IMP to AMP (Nash and Henderson, 1982).

1.5.2.2 Adenosine requiring mutants

Adenine/adenosine requiring auxotrophs have been isolated at five loci, *ade1*, *ade2*, *ade3*, *ade4* and *ade5* (Falk, 1973; Falk and Nash, 1974b; Naguib, 1976; Naguib and Nash, 1976; Johnstone *et al.*, 1985; Nash, personal communication). The first one, *ade1*, maps on the X chromosome. This was a weak, slow growing mutant. Because the sole *ade1* mutant has been lost, no further study has been done.

Both *ade2* and *ade3* are located on the second chromosome. The prototype *ade3* mutant, *ade3*^{1}, has no obvious visible phenotype, but *ade2* shows a mutated dull red eye-colour, a phenotype that implies a pteridine deficiency. Supplementation with adenosine, adenine and inosine can successfully rescue them but hypoxanthine, guanine and guanosine cannot (Naguib, 1976; Johnstone *et al.*, 1985). Both of them have been cloned. Sequence analyses have revealed that *ade2* encodes FGARAT (Tiong and Nash, 1993) and *ade3* encodes the trifunctional enzyme GARS/GART/AIRS (Henikoff *et al.*, 1986b).

Both *ade4* and *ade5* are located on the X chromosome. These mutants are leaky mutants. In addition to adenine sources, the *ade4* and *ade5* auxotrophs can be partially rescued by guanine sources (Nash, personal communication). *In situ* hybridization revealed that the AIRC-SAICARS gene is located in the same region (O'Donnell and Clark, 1997).

The mutants of *Prat* which encodes the first enzyme of *de novo* IMP biosynthesis should behave similarly to the other *ade* mutants even though they have not been tested.

1.5.2.3 Guanosine requiring mutants

Two guanosine auxotrophs have been isolated, *gua1*^{ts} and *bur*$^{gua2-1}$. The *gua1*^{ts} map to the X chromosome in the cytological band *9E1-9E4* (Falk, 1973; Falk and Nash, 1974b; Johnson *et al.*, 1979; Nash *et al.*, 1981). This mutant is a leaky, temperature sensitive auxotroph showing about 5% viability at 29°C and 50% viability at 25°C without guanosine supplementation.

The second guanosine auxotroph *bur*$^{gua2-1}$ is an allele of the eye-colour mutant *burgundy* (*bur*) located on the second chromosome (Johnstone *et al.*, 1985). It absolutely requires guanosine supplementation and is the strongest mutation among all purine auxotrophs (Johnstone and Nash, 1979). It shows a mutant eye-colour similar to *ade2* or *ras*, implying a deficiency of pteridine levels.

Adenosine fails to supplement both guanosine mutants. An auxotroph that can only be rescued by exogenous guanosine is probably defective in biosynthesis from IMP to GMP. There are two enzymes involved in this reaction, IMP dehydrogenase and GMP synthetase. Further analyses of these mutants have revealed that IMP dehydrogenase is encoded within a genetically complex locus that includes the *gua1* locus (the subject of this thesis) and *bur* locus encodes GMP synthetase (Nash *et al.*, 1994; Chomey and Nash, 1995).

1.5.2.4 Purine requiring mutants

Falk and Nash (1974b) isolated two mutants which can be supplemented by either adenosine or guanosine. These auxotrophs, which are allelic, were assigned to a locus named *pur1* and mapped at approximately the same location as *gua1*^{ts} on the X chromosome (Johnstone and Nash, 1979). Both *pur1* alleles have normal eye-colour and minor wing shape defects. The two alleles respond differently to exogenous purine sources; *pur1*^{2} responds equally well to both adenosine and guanosine supplementation, whereas *pur1*^{1} responds less well to adenosine under the same condition (Falk and Nash, 1974b).

1.5.3 The *raspberry* complex

1.5.3.1 *raspberry* eye-colour mutants

The classical red-brown eye-colour mutation *raspberry (ras)* maps at approximately the same location as *pur1* and *gua1* (*9E1-9E4*, X chromosome). It was first discovered by Muller in 1928 and contains four spontaneous eye-colour mutant alleles, *ras*^{1}, *ras*^{2}, *ras*^{3} and *ras*^{4} (Dunn, 1934; Muller, 1935; Beadle and Ephrussi, 1936). These *ras* mutants are defective in pteridine biosynthesis and the mutant eye-colour is caused by a reduction of orange-red drosopterin eye pigments (Nolte, 1959). They are not auxotrophs (Janca *et al.*, 1986). Evans and Howells (1978) found that the GTP cyclohydrolase activity is significantly reduced in the heads of young *ras* adult flies, although at normal or slightly elevated levels at some other stages of the life cycle. They speculated that *ras* may play a role in the developmental regulation of GTP cyclohydrolase. Numerous mutations have been induced at the *ras* locus and almost all of them are lethal mutants *(ras-l)*. This suggests that the *ras* locus is essential for Drosophila (Johnson *et al.*, 1979; Nash *et al.*, 1981; Nash and Janca, 1983).

1.5.3.2 The relationship of *gua1*, *pur1* and *ras*

The three purine metabolism related mutations, *ras*, *gua1* and *pur1* all map to the same cytological location, (*9E1-9E4*) on the X chromosome. Because they complement each other with respect to auxotrophic phenotype and eye-colour, they were initially defined as three individual genes (Johnson *et al.*, 1979, Nash *et al.*, 1981). However, lethal alleles of *raspberry*, *ras-l* mutants, are often allelic to *pur1* and *gua1* mutants, as well. This indicates that these three apparently different loci are functionally related to each other in some way, since the complementation pattern suggests that these three functions are only quasi-independent. Thus the three loci that had previously been considered independent were grouped together under the collective name of the "*gua1-pur1-ras* complex" (Johnson *et al.*, 1979; Nash *et al.*, 1981).

To understand the function of the *gua1-pur1-ras* complex, 24 EMS induced recessive lethal mutations in the region *9E1-9E13* on the X chromosome were selected to conduct a systematic complementation analysis (Janca *et al.*, 1986). Among them, eight of particular interest fell into a single lethal complementation group. Six members of this group failed to complement any mutants of the *gua1-pur1-ras* complex regarding the auxotrophic phenotype or *ras* eye-colour. The remaining two failed to complement *gua1* and *pur1* mutants, but complemented *ras*, producing wild-type eye-colour in the relevant heteroallelic combinations. This result indicates that there may be two interrelated functions in *gua1-pur1-ras* complex. One is involved in eye-colour formation and the other is related to the vital function. The lesion in *raspberry* is not necessarily related to the essential function and the lethality of *ras-l* mutants is probably due to a *pur1* or *gua1* deficiency but not obligately associated with the production of eye pigmentation (Janca *et al.*, 1986).

1.5.3.3 Molecular analysis of the *ras* locus

To further investigate the function of the *gua1-pur1-ras* complex, Leonard (1986) generated five P-element induced *ras-l* mutants, *ras*$^{l22\text{-}l26}$. *In situ* hybridization showed that P-element material was inserted in the *ras* locus (*9E1-9E4*) of four of them, *ras*$^{l23\text{-}l26}$. The *ras*^{l24} strain was chosen to make a genomic _ EMBL-4 library and one phage (LF1) containing*ras* DNA was isolated on the basis of the transposon tag present in the *ras*^{l24} mutation (Leonard, 1986).

This thesis is a continuation of the investigation of the *ras* locus (the *gua1-pur1-ras* complex), that was started more than 25 years ago. The thesis is focused on using the molecular technology to further analyze the function and regulation of the *ras* gene. During the thesis work, the wild-type gene was cloned. The gene structure was deduced from analyzing the genomic and cDNA sequences. Conceptual translation of the longest cDNA revealed that IMP dehydrogenase, the rate limiting enzyme in GMP biosynthesis, is encoded at the *ras* locus. Alterations of IMPDH activity can explain different *ras* mutants, as discussed in the thesis. Several DNA regulatory elements were found by either the comparison of them to known regulatory sequences or the analysis of the molecular lesions identified in the four *ras* eye-colour mutants. The developmental profiles of the concentration of IMPDH transcript and the activity of IMPDH in both wild-type and *ras* eye-colour mutant flies were also studied.

2 Materials and Methods

2.1 Stocks

2.1.1 *Drosophila melanogaster* stocks

D. melanogaster variants used in this study are described in (Table 2-1).

Am^+ was used as the standard wild-type strain in our laboratory and the original samples were collected from Am^+. During embryo collection, it was found that a large proportion of eggs collected failed to produce larvae. This implies that the Am^+ females produce a large proportion of unfertilized eggs or inviable embryos. For this reason, the wild-type strain Oregon R (OrR), from which the Am^+ was derived, was utilized for the embryonic and early larval samples.

The late larval and pupal samples had been sexed and collected from Am^+ at the time when Am^+ inviability was discovered. Since re-harvesting of sexed samples would have taken more than two person-months, it was decided to utilize these samples. As a precaution, it was checked that the survivorship of late third instar larvae to adult stage is similar (>90%) in both OrR and Am^+. Northern analysis also demonstrated no obvious difference between the transcript patterns in viable late third instar larvae samples of OrR and Am^+ flies (Figure 3-23). In consequence, these sexed late larval and pupal samples from Am^+ were used in this study.

2.1.2 Bacteria, bacterial plasmid vectors and bacteriophage λ vectors

2.1.2.1 Bacteria:

The bacterial host cells are derived from *E. coli* and are required for different cloning vectors (Table 2-2).

2.1.2.2 Bacterial plasmid vector pUC19

pUC19 (Vieira and Messing, 1982; Yanisch-Perron *et al.*, 1985) is a high copy number plasmid vector. It contains a fragment of the lacZ (β-galactosidase gene) DNA which encodes the N-terminal portion of β-galactosidase and can complement lacZΔM15, a defective form of the lacZ gene encoding the C-terminal portion of the β-galactosidase gene. LacZ expression can be induced by isopropylthio-β-D-galactoside (IPTG). When pUC19 is introduced into a gal^- host bearing lacZΔM15, functional β-galactosidase will be produced and form blue colonies if plated on media containing 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and IPTG. The cloning site (polylinker) of pUC19 is located inside the lacZ fragment, insertion of a foreign DNA into pUC19 will usually destroy the lacZ function and produce white colonies. Thus, the insert bearing pUC19 can be distinguished by the colour (blue/white selection).

Table 2-1 Description of *D. melanogaster* stocks

Stock	Description	Source
Wild-type		
Oregon R	Derived from wild flies	Roseburg, OR. USA
Amherst	Inbred derivative of Oregon R	Amherst College
ras eye-colour mutants		
ras^{1}	*ras*^{1} *dy*^{1}	Caltech Stock Center
ras^{2}	*ras*^{2}	Caltech Stock Center
ras^{3}	*ras*^{3} *m*^{1}	Caltech Stock Center
ras^{4}	C(1)DX, *y*^{1} *f*^{1}/*lz*^{BS} *lz*^{46} *ras*^{4} *v*^{1}	Mid-America Stock Center (Bowling Green)
P-element induced *ras* lethal mutants		
ras^{l22}	*y cv ras*^{1} *v f* /FM6	N. J. Leonard
ras^{l23}	*y cv ras*^{1} *v f* /FM6	N. J. Leonard
ras^{l24}	*y cv ras*^{1} *v f* /FM6	N. J. Leonard
ras^{l25}	*y cv ras*^{1} *v f* /FM6	N. J. Leonard
ras^{l26}	*y cv ras*^{1} *v f* /FM6	N. J. Leonard

Table 2-2 Bacteria and the cloning vectors

Host	Vector	Growth Medium	Selection
Q358	λ phage (EMBL-4)	LB/MgSO4	None
NM538	λ phage (EMBL-4)	LB/$MgSO_4$	None
Q359	λ phage (EMBL-4)	LB/$MgSO_4$	Spi^-/P2
NM539	λ phage (EMBL-4)	LB/$MgSO_4$	Spi^-/P2
ECN71	pUC19	LB/ampicillin	blue/white
XL1-Blue	pUC19/pBluescript	LB/ampicillin	blue/white

2.1.2.3 Bacterial phagemid vector pBluescript

pBluescript phagemid (Stratagene) is actually a pUC plasmid with an M13 (either M13+ or M13-) phage origin so it can grow as either a plasmid or a phage. In addition to its pUC plasmid's merits, high copy number and blue/white selection, pBluescript has the following advantages.

- a) As a phage, it can synthesize single strand DNA from either strand depending on the M13+ or M13- origin.
- b) T3 and T7 promoters are placed adjacent to either end of the polycloning site (polylinker) so the inserts can be transcribed *in vitro* from either direction.
- c) The polylinker is bounded by restriction enzyme *Sac* I and *Kpn* I sites and can be placed in either direction. Both *Sac* I and *Kpn* I generate 3′ overhanging ends. This property is useful in generating Exo III nested deletions.
- d) There are four kinds of pBluescripts depending on the M13 origin and the orientation of the *Sac* I and *Kpn* I sites, SK(+), SK(-), KS(+) and KS(-). A given insert in any form can be easily switched into another form. Thus, it is very convenient to generate single strand DNA from either strand and nested deletion series from either orientation.

2.1.2.4 Bacterial phage vector EMBL-4

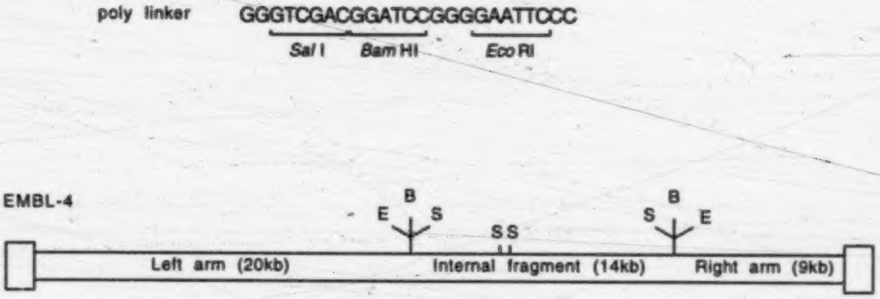
The replacement bacteriophage vector λ EMBL-4 (Frischauf, *et al.*, 1987) was used to make a wild-type genomic library. EMBL-4 is designed to take the advantage of λ phage lysogens. The intact EMBL-4 cannot grow in lysogenic host (NM539 or Q359) carrying prophage P2 (sensitive to P2 interference, Spi^+) because the internal replacing fragment contains red and gam genes which confer a Spi^+ phenotype. When the internal fragment is replaced by foreign DNA, the recombinant is Spi^- and thus can grow in a P2 lysogenic host. There are three restriction sites (*Eco* RI, *Bam* HI and *Sal* I) at each end of the replacement fragment (Figure 2-1) and, in the normal cloning practice for which the vector is designed, the *Bam* HI sites generate the boundaries of the insert. This is particularly useful because it accepts *Sau* 3A fragments produced in partial digests which contain the same GATC cohesive ends as *Bam* HI (*Section 2.4.7.2.*). The *Eco* RI site is external to the *Bam* HI site and can be used to excise the insert from recombinant phage. *Sal* I is internal to the *Bam* HI site and used to eliminate the *Bam* HI ends of the internal fragment and to fragment it (It contains four *Sal* I sites) thus ensuring that the internal fragment will not religate to the *Bam*-HI cohesive ends of the vector arms.

2.2 Media and solutions

2.2.1 Drosophila medium

All media and solutions were prepared using chemicals from Difco Laboratories, Sigma, Bethesda Research Laboratories (BRL), Boehringer Mannheim or Fisher Scientific. The solutions were made in distilled/deionized water and autoclaved at 121°C for 15 minutes unless otherwise stated.

Figure 2-1 Restriction map of λ EMBL-4



Diagrammed is a restriction map of λ EMBL-4 used as the genomic cloning vector in this study. The intact EMBL-4 cannot grow in lysogenic host (NM539 or Q359) because the internal fragment contains red and gam genes which confer a Spi$^+$ phenotype. When the internal fragment was replaced by genomic DNA (range from about 10-25kb), the recombinant becomes Spi$^-$ and thus can grow in the lysogenic host.

Two polylinkers are located at each end of the internal fragment. There are three restriction sites (*Eco* RI, *Bam*-HI and *Sal* I) at each polylinker but on the opposite directions. These restrictions sites are useful in cloning. *Bam* HI (B) site is particularly useful to clone the *Sau* 3A partially digested genomic DNA fragments. *Eco* RI (E) can be used to excise the insert from the recombinant phage. *Sal* I (S) can be used to eliminate the religation of the internal fragment to the vector arms.

All Drosophila stocks were grown on standard yeast-sugar medium composed of 100g Brewer's yeast, 100g sugar, 10g granulated Bacto-Agar, 2.7g $NaH_2PO_4 \cdot H_2O$ and 4.3g Na_2HPO_4 per 1,000ml distilled water. The medium was boiled gently in a steam kettle for 40-45 minutes or autoclaved at 110°C for 45 minutes. After the medium cooled to 70°C, propionic acid (10ml/L) and tetracycline (10mg/L) or chloramphenicol (1g/L) were added and mixed thoroughly before pouring (modified formula from Nash and Bell, 1968).

2.2.2 Bacterial media

2.2.2.1 Luria-Bertani (LB) Broth:

LB consists of 10g Bacto-Tryptone, 5g Bacto yeast extract and 10g NaCl per liter (pH7.5). If necessary, antibiotic was added to the medium just before use.

LB plates were prepared by adding 15g Bacto-agar to one liter LB broth and autoclaved before addition of antibiotic (if needed) and pouring.

LB top agarose was prepared by adding 0.7g low melting-point agarose to per 100ml LB broth (in small bottles) and autoclaved. The LB top agarose was remelted in a microwave oven before use.

2.2.2.2 Tryptone (T) Medium:

T broth is nutritionally less rich than LB broth. It was mainly used to generate bacteriophage plate lysates. It contains 10g Bacto-Tryptone and 5g NaCl per liter (pH7.5).

T plates were prepared by adding 15g Bacto-agar to per liter T broth and autoclaved before pouring.

T top agarose was prepared similarly to LB top agarose.

In the preparation of bacteriophage, 10mM $MgSO_4$ was added to the LB or T agar medium before autoclaving and 10mM $MgSO_4$ and 0.2% maltose were added to top agarose before plating.

2.2.2.3 2YT MEDIUM:

2YT is nutritionally richer than LB. It was used in single-strand DNA (ssDNA) preparation for use as sequencing templates. 2YT consists of 16g Bacto-Tryptone, 10g Bacto yeast extract and 5g NaCl per liter (pH7.0). The solution was distributed into 100ml bottles and autoclaved.

2.2.3 Stock reagent solutions

These are the complex reagents; simple solutions (containing just one solute) are described in the text.

100X Denhardt's reagent	2.0% (w/v) Ficoll (400,000MW), 2.0% (w/v) Polyvinylpyrrolidone (360,000MW), 2.0% (w/v) Bovine Serum Albumin (Keep in small aliquots at -30°C after filter sterilization)
5X PE	0.5% (w/v) sodium pyrophosphate, 5.0% (w/v) SDS, 1.0% (w/v) Polyvinylpyrrolidone (400,000MW), 1.0% (w/v) Ficoll (400,000MW), 1.0% (w/v) BSA, 25mM EDTA, 250mM Tris·HCl (pH7.5) (Filter sterilization, keep at room temperature)
6X loading buffer	0.25% bromophenol blue, 0.25% xylene cyanol FF, 25% ficoll (type 400)
20X SSC	3M Sodium Chloride, 0.3M Sodium Citrate
20X SSPE	3.6M Sodium Chloride, 0.02M EDTA, 0.2M Sodium Phosphate buffer (pH7.7)
RNA extraction buffer	4M guanidinium thiocyanate, 25mM pH7.0 sodium citrate, 0.5% sarcosyl, 0.72% _-mercaptoethanol (No sterilization, keep at room temperature up to three month)
SM	100mM Sodium Chloride, 8mM Magnesium Sulfate, 50mM Tris·HCl (pH7.5)
TBE buffer	89mM Tris·HCl, 89mM boric acid, 2mM EDTA
TE	10mM Tris, 1mM EDTA (pH8.0)

2.3 Tissue collection

2.3.1 Embryos and earlier larval samples

All fly cultures used for harvesting tissue were maintained at 25°C. The tissue samples were collected at room temperature. The samples were frozen in liquid nitrogen and kept at -75°C until use.

In this study, embryos, larvae, pupae and young imagos were collected from stocks homozygous for ras^{1-3} and from Oregon R (or its inbred derivative, Amherst). When sexing was possible, the male and female samples were collected separately. Since ras^{4} is female sterile, only sexed ras^{4} young imagos were collected.

The embryos and earlier larval samples were collected as the following:

Several thousand 2-3 day old flies were placed in a population cage with 2 or 3 150mm petri dishes containing standard yeast/sucrose medium coated with yeast paste. After a predefined egg-laying period, the plates were taken out of the cage and left at 25°C to different developmental stages (Table 2-3). The egg or larval samples were then washed from the petri dishes onto a fine mesh screen under cold running water, with the aid of gentle brushing, and rinsed until the yeast paste was removed completely. Approximately 50 mg samples were distributed into Eppendorf tubes and frozen.

2.3.2 Late larvae and pupae

Late third instar larvae were sexed under the microscope (with the extensive assistance of David Nash). Around 45-50 larvae of a given sex (which is enough for one RNA or protein extraction) were put into Eppendorf tubes and frozen.

For collecting pupal samples, massed samples of sexed larvae were put into fresh medium containing culture tubes and incubated at 25°C. The purpose of the medium was to provide appropriate humidity levels to allow continuation of development. Pupal samples were collected after approximately 24, 48, 72 and 96 hours. Samples of about 40-50 sexed pupae were put into Eppendorf tubes and frozen.

2.3.3 Adults

For DNA extractions, 0-24 hour old imagos were used. For RNA and protein extractions 0-8 hour old imagos were collected. Males and females were collected separately. Samples of fifty imagos were put into Eppendorf tubes and frozen.

Table 2-3 Protocol for collection of embryos and early larvae

Sample name	Sample age (hours)	Egg lay period (hours)	Incubate at 25°C (hours)
0-2 hours	0-2	0-2	0
2-4 hours	2-4	0-2	2
4-8 hours	4-8	0-4	4
8-16 hours	8-16	0-8	8
16-20 hours	16-20	0-4	16
20-40 hours	20-40	0-20	20
40-64 hours	40-64	0-24	40
64-88 hours	64-88	0-24	64

2.4 DNA manipulation

2.4.1 DNA extraction

2.4.1.1 Drosophila genomic DNA

Drosophila genomic DNA extraction was carried out using a protocol described by Ish-Horowicz, *et al.* (1979), as modified by W. Clark, University of Alberta (personal communication). To avoid shearing the DNA, special precautions were applied: All manipulations were done very gently; samples were never vortexed and plastic pipette tips were cut to make them wider.

50-100 frozen sexed, young adult flies (about 50-100mg) were transferred to a 7ml Dounce homogenizer with 1.5ml of fresh made 10mM Tris (pH7.5), 60mM EDTA, 0.15mM spermidine, 0.2mg/ml pronase E (type XXV) and 50µg/ml RNase A. The samples were ground using 4-5 slow strokes. The homogenate was incubated at room temperature for 30 minutes and inverted slowly every 10 minutes. Then, 1.5ml of freshly made 0.2M Tris (pH9.0), 30mM EDTA, 2% SDS and 0.2mg/ml pronase E (type XXV) were added to the samples and mixed well. The samples were then incubated at 37°C for 90 minutes. The digest was extracted twice with phenol:chloroform (1:1) mixture and once with chloroform using a bench shaker, for 30 minutes each time. After chloroform extraction, the supernatant was transferred to a new tube, 5M NaCl was added to a final concentration of 0.2M and the DNA was precipitated with ethanol at -70°C overnight. The pellet was washed once with 70% ethanol, vacuum dried and resuspended in 200µl TE.

2.4.1.2 Large scale plasmid DNA purification

Large-scale plasmid DNA isolation employed a protocol of D. Nash (personal communication): A single plasmid-containing bacterial colony was cultured in 5ml LB containing 100µg/ml ampicillin at 37°C, overnight. This culture was used to start a new overnight culture in 400ml LB with the same antibiotic regime. The larger overnight bacterial culture was centrifuged in a 500ml bottle at 5000RPM in a Sorvall GSA rotor at 4°C for 10 minutes. The pellet was resuspended in 8ml of 50mM Tris (pH8.0) with 25% sucrose on ice, after which 2.6ml of 10mg/ml lysozyme in 10mM Tris (pH8.0) was added to the suspension. The mixture was swirled very slowly on ice for 10 minutes and 13.2ml of lysis solution (50mM Tris, 62.5mM EDTA, 0.1% Triton X-100, pH8.0) added. The mixture was again swirled on ice for 10 minutes. At this stage, the cells should are lysed and the DNA released making the solution very viscous. This viscous solution was transferred to a 30ml plastic tube and centrifuged at 15,000RPM for 50 minutes at 4°C in a Sorvall SS-34 rotor to sediment bacterial debris. The supernatant was carefully removed to a new tube, taking care not to disturb the gelatinous pellet, which contains most of the bacterial genomic DNA.

Purification of the circular plasmid DNA after bacterial lysis was carried out using equilibrium centrifugation in cesium chloride gradients (Sambrook, *et al.*, 1989). CsCl was added to the lysate to a final concentration of 1g/ml. The solution was shaken gently in a 37°C water bath until the CsCl was totally dissolved. Ethidium bromide was added to the DNA/CsCl solution to a final concentration of

740µg/ml (0.8ml of 10mg/ml Ethidium bromide in every 10ml DNA/CsCl solution). The solution was mixed well and centrifuged at 8,000RPM for 10 minutes in a Sorvall SS34 rotor to bring the frothy scum which consists of complexes of Ethidium bromide and bacterial protein to the surface. The clear red solution under the scum was transferred to Beckman Quick-Seal tubes and centrifuged at 45,000RPM for 16 hours in a VTi65 rotor. After centrifugation, the circular DNA was collected, the Ethidium bromide was removed with CsCl-saturated isopropanol, and the DNA was EtOH precipitated. The purified DNA was assessed electrophoretically and the DNA concentration was measured at A_{260}.

2.4.1.3 Small scale plasmid and phage DNA extraction

Small scale plasmid DNA extraction was done using the "Alkali Lysis" method and small scale phage DNA extraction was done using the "Plate Lysate" method, both described in Sambrook *et al.* (1989) except that deproteinization employed two phenol/chloroform extractions, followed by one chloroform extraction.

2.4.1.4 Single strand phagemid DNA (ssDNA) extraction

Single-stranded phagemid DNA for use as sequencing template was produced using the polyethylene glycol (PEG) precipitation method of D. Nash (personal communication): XL1-blue bacteria containing appropriate pBluescript-derived DNAs were grown on LB plates containing ampicillin (100µg/ml) and tetracycline (12.5µg/ml). A single colony was inoculated into 2.5ml 2YT broth with 100µg/ml ampicillin, 0.001% thiamine and 10^7 pfu/ml VCSM13 helper phage and incubated at 37°C for 90 minutes. Then kanamycin A was added to a final concentration of 70µg/ml and the mixture was incubated at 37°C for another 16-20 hours, until it became cloudy. The culture, now containing a mixture of bacteria and phagemids, was centrifuged and 1.8ml supernatant was transferred to a new 2.2ml Eppendorf tube. To precipitate the phagemid particles, 450µl of 20% PEG 8000 in 2.5M NaCl was added to the supernatant and mixed well. The tube was left at room temperature for 15 minutes then centrifuged in an Eppendorf micro-centrifuge at full speed for 15 minutes. The phage pellet was resuspended in 300µl TE followed by phenol/chloroform and chloroform extraction and ethanol precipitation. The ssDNA pellet was dissolved in 20µl TE. The quality and quantity were assessed by running 1µl of the DNA sample on an agarose mini-gel.

2.4.2 Restriction endonuclease digestion

Restriction endonuclease digestions were performed according to each enzyme manufacturer's suggestions with the buffers supplied. In the case of double enzyme digestions, unless the two enzymes used the same kind of buffer, the two digestions were performed sequentially, separated by an ethanol precipitation to allow the second enzyme digestion to be carried out under the optimal conditions.

In the case of partial *Sau* 3A digestions, the initial activity in the available commercial enzyme preparation was too high to allow accurately repeated small-scale dilution on a digestion by digestion basis. In order to have better control of the reaction, an appropriately large quantity of the commercial preparation was diluted in the manufacturer's recommended storage buffer (10mM Tris(pH7.4), 50mM

KCl, 0.1mM EDTA, 1mM DTT, 500μg/ml BSA and 50% v/v glycerol). The activity of this diluted sample on Drosophila genomic DNA was established and the same sample was then used in all subsequent digestions.

2.4.3 DNA agarose gel electrophoresis

Digested DNAs were separated in 0.4-1% agarose gels in TBE buffer including 0.25μg/ml Ethidium bromide. A 6X loading buffer was added to the samples before loading. BRL 1kb ladder was normally used as a size standard.

2.4.4 Gel Purification

Digested DNA fragments were run on 1.0% TBE agarose gels as above, but without Ethidium bromide. DNA ladder was loaded on the same gel with 0.25μg/ml Ethidium bromide in the 6X loading buffer. Desired DNA fragments were excised under UV light from gels according to the size markers. The excised agarose blocks were positioned on top of a glass wool plug in the bottom of a pierced 0.5ml Eppendorf tube and frozen in liquid N_2 for five minutes. The Eppendorf tube was put into a larger Eppendorf tube and centrifuged for 15 minutes at full speed. DNA containing liquid extruded from the agarose gel collected in the larger tube and the DNA was ethanol precipitated from it.

2.4.5 Preparation of radiolabeled DNA probes

The "Multiprime" DNA labeling system (Amersham) was used to make DNA probes for Southern blots, genomic library screens and for some northern blots. These probes were purified by passage through a Sephadex G-50 column (Sambrook *et al.*, 1989). The radioactivity of labeled probe was estimated from an aliquot of the purified probe by scintillation spectrometry. Generally, 10^5 cpm (counts per minute) probe/ml hybridization solution were used in Southern hybridization and genomic library screen and 10^6 cpm/ml were used in northern hybridization.

2.4.6 Southern transfers and hybridization

DNA samples were digested by appropriate enzymes and electrophoresed through agarose TBE gels. The gel was exposed under UV light to nick the DNA then denatured in 0.5M NaOH and 1.5M NaCl for 30 minutes. The gel was then neutralized in 0.5M Tris/HCl (pH7.4) and 1.5M NaCl for 90 minutes. BIOTRANS Nylon Membrane (ICN Biomedicals, Inc.) was used in Southern transfer (see manufacturer's manual). The capillary blot transfers were carried out in 20X SSC and the transferred DNA was fixed to the membrane by baking at 80°C for 15 minutes followed by UV irradiation for about 15-20 seconds on UV transilluminator. Hybridizations were performed overnight in 5X Denhardt's Solution, 5X SSC, 1% (w/v) SDS with 100μg/ml denatured Salmon sperm DNA, at 65°C in a water bath, as described by the manufacturer's manual supplied with the membrane. After hybridization, the membrane was washed in 2X SSC, 0.1% SDS for 5 minutes at room temperature twice followed by three times in 0.1X SSC, 0.1% SDS at 50°C for 15 minutes each. Autoradiography was performed at -70°C using Fuji RX Medical X-ray film.

2.4.7 Making and screening an EMBL-4 genomic DNA library

2.4.7.1 Preparation of vector arms

A sample of the replacement bacteriophage vector λ EMBL-4 prepared by N. J. Leonard was used to make the genomic library. EMBL-4 DNA was first digested with *Bam*-HI then digested with *Sal* I. Ethanol precipitation was performed after each digestion. *Bam* HI digestion produces three DNA fragments, a left arm, a right arm and a 14kb internal fragment, which is non-essential to the lytic cycle of the λ bacteriophage from which EMBL-4 is derived and can be replaced by exogenous DNA fragments from 15-25kb in length. *Sal* I digestion abolishes the capacity of the internal fragment to be religated to the arms during library construction. It also cuts the fragment into five pieces, facilitating confirmation, at the level ascertainable by gel electrophoresis, that the digestion is complete.

2.4.7.2 Preparation of genomic inserts

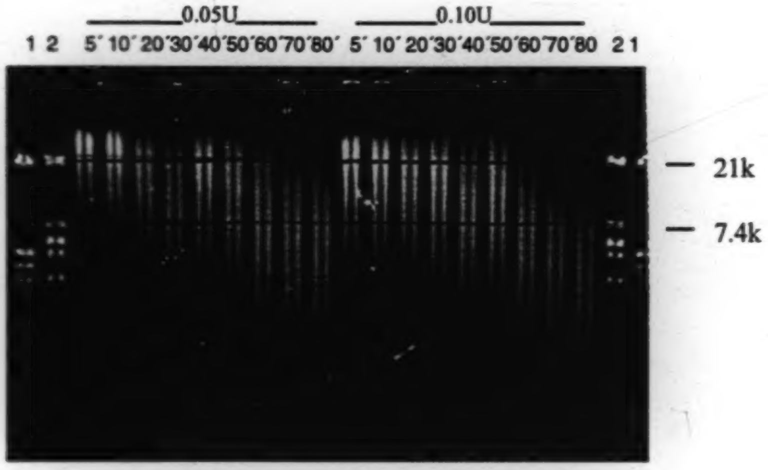
DNA from wild-type Am^+ embryos, 0-18 hours old, was used to make a genomic library. The DNA samples had been purified by N. J. Leonard (1986) several years previously, so they were checked on a 0.4% agarose gel before use. Samples appearing to be of high molecular weight and without obvious sign of degradation were selected. To get quasi-randomly cut DNA fragments, the DNA was partially digested with *Sau* 3A, a restriction enzyme with a four base pair recognition site. Four base pair cutters cut DNA with an average spacing of 4^4(=256) base pairs in a random 50% GC DNA sequence. With a four-base cutter, partial digestion products of sizes suitable for insertion into EMBL-4 (15-25kb) and including most, if not all, sequences in any genome might be expected. Only unusually long sequences that are, for some reason, impoverished for the actual target sequence (in this case GATC) would be excluded. *Sau* 3A produces the same four base cohesive ends (GATC) as *Bam* HI, which was used to prepare the vector arms, so that its restriction products are directly ligatable to the arms.

Sau 3A enzyme recognition sites are not necessarily evenly spread along the genomic DNA sequence. Consequently, in order to include DNA sequences that are both rich and poor in *Sau* 3A sites within the panel of partial digestion products, digestions were carried under a wide range of conditions: Digestions were made with 0.05 and 0.10 Units of *Sau* 3A/µg DNA at 37°C; samples were taken at 5, 10, 20, 30, 40, 50, 60, 70 and 80 minutes. Aliquots from each sample were checked electrophoretically (Figure 2-2).

All the *Sau* 3A digestion products were mixed and size-separated by NaCl gradient (5-29% NaCl in TE, pH8.0). A 200µl sample (100µl *Sau* 3A digested DNA and 100µl 5% NaCl) was loaded on the top of a pre-prepared 5-29% NaCl gradient in an SW40 tube and centrifuged at 35K, 22°C, 6 hours in an SW40 rotor. The centrifuged samples were collected in about 300-500µl fractions and sized on a 0.4% agarose gel. λ DNA digested with *Bgl* II was loaded with the samples as standards. The samples containing DNA of appropriate sizes (15-25kb) were desalted by dialysis (20mM Tris pH7.5, 1mM EDTA, 1.5 hour 3 times), concentrated (Elutip-d, Schleicher & Schuell) and ethanol precipitated.

Figure 2-2 ***Sau*** **3A partially digested Drosophila genomic DNA**

A



The digestions were made with 0.05U and 0.10U of *Sau* 3A/µg Drosophila genomic DNA. Samples were taken at various timing points.

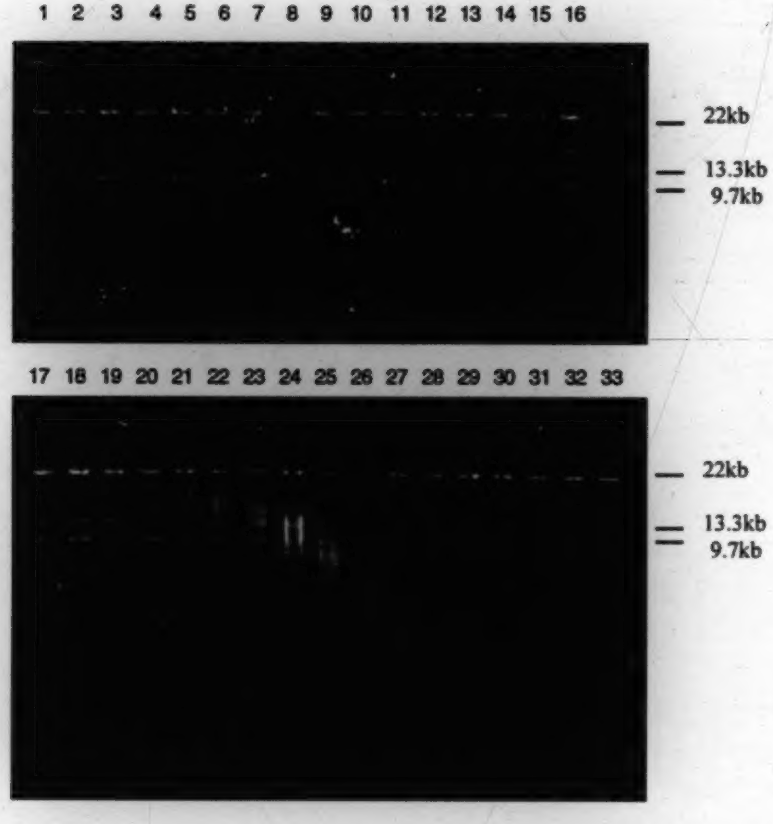
Size marker

1. *Eco* RI and *Hin* dIII digested _ DNA
2. *Eco* RI digested _ DNA

Sau 3A digested Drosophila DNA, samples taken at

5′= 5 minutes
10′= 10 minutes
20′= 20 minutes
30′= 30 minutes
40′= 40 minutes
50′= 50 minutes
60′= 60 minutes
70′= 70 minutes
80′= 80 minutes

B



NaCl separated *Sau* 3A partially digested Drosophila DNA fragments. The background size marker is *Bgl* II digested λ DNA. The purified DNA fragments in samples 7 to 16 are mainly larger than 22kb and the DNA fragments in samples 24 to 30 are mainly smaller than 13kb. Samples 18, 19, 20, 21 and 22 were mixed and used to make the genomic DNA library.

2.4.7.3 Ligation of the arms and inserts

Ligation was carried out overnight at 15°C in a 20μl reaction (66mM Tris(pH7.5), 10mM $MgCl_2$, 5mM DTT and 1mM ATP) with 0.2U T4 DNA ligase (BRL). A range of arms:inserts ratio were tested using a fixed EMBL-4 arms concentration (0.2μg/reaction) and various amounts of *Sau*-3A partial digested DNA at 1:4, 1:2, 1:1, 1:0.5 and 1:0.25. The 1:1 and 1:2 ratios were chosen for the final bulk ligations, in which the reaction volume was 40μl. Ligation products were packaged *in vitro* using Packagene Extract System (Promega).

2.4.7.4 Library size

In preparing a genomic DNA library, the number of recombinant clones required to carry effectively all fragments of a given genome must be calculated. A formula described by Sambrook, *et al.* (1989) was used in this study. The formula is based on an average size of the inserts and on the size of the genome concerned. To increase the probability of having all genomic regions included in the library, and because of the partially digested overlapping genomic DNA fragments, the required clone number is much higher than the number of the same size nonoverlapping fragments in the same genome. The formula is as follows

$$N = \ln(1 - P)/\ln(1 - f),$$

N = the number of recombinants required,
P = the probability of recovering a given sequence,
f = the fraction of the genome found in an average single recombinant,

Assuming the average size of the insert is 20kb and the genome size of *Drosophila melanogaster* is $1.65\text{x}10^8$bp, then

$$f = (2\text{x}10^4)/(1.65\text{x}10^8)$$

With 99% probability (P = 0.99), the number of recombinants needed in the library is calculated to be 37,990.

$$N = \ln(1 - 0.99)/\ln(1 - 2.0\text{x}10^4/1.65\text{x}10^8)$$
$$N = 37{,}990$$

With 99.99% probability (P = 0.9999), the number is 75,981.

$$N = \ln(1 - 0.9999)/\ln(1 - 2.0\text{ x}10^4/1.65\text{x}10^8)$$
$$N = 75{,}981$$

Approximately 38,000-76,000 recombinants are needed for a 99-99.99% complete *Drosophila melanogaster* library of about 20kb sized inserts.

2.4.7.5 Library amplification

A sample of the packaged material was plated on the P2 lysogen NM539 on a 90mm plate overnight to test the efficacy of ligation and packaging. The high titer samples were then plated at approximately 10^4 plaques per 150mm plate. A total of about 400,000 plaques of the primary library was generated in the initial plating. This once amplified library was harvested by adding 20ml SM to each plate and swirling slowly at 4°C for several hours to release the phage particles. The phage-suspension was collected, chloroform was added to a final concentration of 0.3% and the amplified library was stored at 4°C.

2.4.7.6 Screening the library

To identify and isolate specific recombinant phage from the library, bacteriophage plaques were screened by hybridization with ^{32}P-labeled probes (Sambrook *et al.*, 1989). Aliquots from the library were plated at around 40,000 plaques per 150mm petri plate and a total of 10 plates was screened. The plaques were transferred to BIOTRANS Nylon Membranes (ICN Biomedicals, Inc.) and probed with ^{32}P-labeled LF1 8.4kb *Eco* RI DNA fragment according to the conditions described above. It should be noted that an M-strain (Am^+) was used to construct the library so that the presence of a segment of a P-element in the probe did not present a problem. Autoradiography of the labeled membranes was performed for 16-48 hours at -70°C using Fuji RX Medical X-ray film.

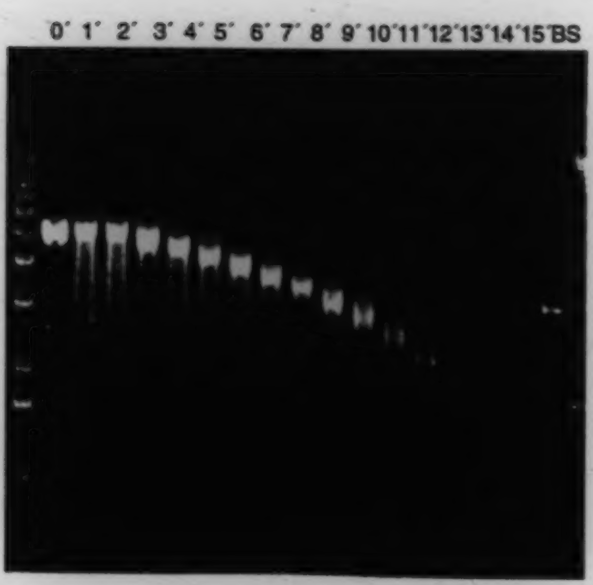
2.4.8 DNA sequencing and computer analysis

The DNA sequencing employed the dideoxy-chain termination method (Sanger *et al.*, 1977). The DNA fragment of interest was subcloned into the M13-derived vector pBluescript SK(+)II or KS(+)I (Stratagene). Nested deletion series of subclones for sequencing both directions were generated using the Exo III/Mung Bean system or the Erase-a-Base system (Promega). Clones differing in size by approximately 200-300bp were selected (Figure 2-3). Single-stranded DNAs (see above *Section 2.4.1.4*) were generated for use as templates. Sequencing reactions utilized the Sequenase Version 2.0 DNA Sequencing kit (United States Biochemical) and ^{35}S-labeled dATP (Amersham). The radioactive sequencing products were separated on a 6% polyacrylamide gel (5.7% acrylamide, 0.3% bis-acrylamide, 8M Urea, 1X TBE). Before pouring, 5µl/ml 10% ammonium persulfate and 0.5µl/ml of the polymerization catalyst TEMED (*N,N,N',N'*-tetramethylethylenediamine) were added. A Model S2 Sequencing Gel Electrophoresis Apparatus (BRL) was employed. After electrophoresis, the gel was fixed in 10% methanol, 7.5% acetic acid for 30 minutes, transferred to Whatman 3MM paper and dried on a gel dryer at 80°C for 60 minutes. Autoradiography was performed for 16-48 hours at -70°C using Fuji RX Medical X-ray film.

The DNA sequence was read manually. Gel compressions, which are normally caused by the secondary structures involving G and C residues, were resolved by substitution of dITP for dGTP in the sequencing reaction. To get accurate sequences, the DNA was sequenced from both directions and the final sequence was determined by comparison of the sequences obtained from both strands.

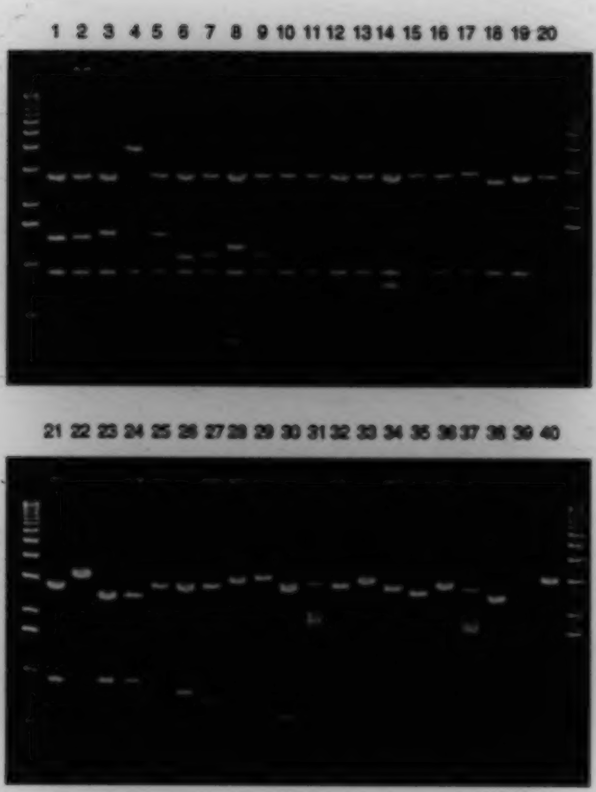
Figure 2-3 Preparation of a nested deletion series of cDNA SK83

A



The cDNA SK83 was cut by *Kpn* I then unidirectionally digested by Exo III/Mung Bean system (25μg *Kpn* I digested DNA, 1000U Exo III, 35°C). Samples were taken at the different times (shown in minutes) at the top of each lane. The partially digested DNA was religated and used to transform XL1-Blue. BS: pBluescript DNA.

B



Photograph of *Bss* HII-digested phagemids (pSK83) recovered from the previous experiment. *Bss* HII cuts the recovered plasmids into three fragments. The longest is about 3kb which contains only the phagemid DNA. The second is about 1.4kb and the third piece is about 0.9kb. The picture shows that the 1.4kb fragment was digested first by Exo III followed by the 0.9kb fragment. The Exo III/Mung Bean system generated a nice 200-300bp deletion series.

The initial DNA sequence assessments used DNA Strider (Marck, 1988; Douglas, 1995). Candidate conceptual translation products were used to search the Swiss Protein library, employing FASTA (Pearson and Lipman, 1988) or the PIR data base using BLAST (Altschul, *et al.*, 1990). Sequences retrieved therefrom were aligned using CLUSTAL V (Higgins and Sharp, 1988, 1989) and Macaw (Multiple Alignment Construction & Analysis Workbench, Karlin and Altschul, 1990; Schuler *et al.*, 1991).

2.4.9 Polymerase Chain Reaction

Amplification of specific segments of genomic DNA by Polymerase Chain Reaction (PCR, Saiki, et al, 1986) employed a method of Dr. K. Roy, University of Alberta (personal communication). The PCR was carried out in a total volume of 100μl with approximately 2μg Drosophila genomic DNA and 0.5μM of each primer; the reaction mixture contained 70mM Tris (pH9.0), 2mM $MgCl_2$, 0.1% Triton X-100, 250μM each of the four dNTPs and 1-2U Taq DNA polymerase which was kindly provided by Gary Ritzel, University of Alberta. Samples were overlaid with 50μl of light mineral oil to prevent evaporation. Thirty cycles of PCR were performed in an Autogene II (Grant Instruments). The initial denaturation was carried out at 94°C for 5 minutes and thereafter for 30 seconds at the same temperature. In each cycle, primer annealing took place at 58°C (30 sec), and extension at 74°C (2 minutes). Primers were synthesized by Dr. Roy.

The PCR products were checked in either 2% agarose or 3% polyacrylamide gels. Appropriate amplified PCR product bands were purified from the gels. The purified DNA fragments were either sequenced directly using the original PCR primers and the double strand DNA method described in Sequenase Version 2.0 DNA Sequencing kit (United States Biochemical) or cloned into pGEM-T vector system I (Promega) for further use.

2.4.10 Oligo primers

The two sets of oligo primers used in this study were prepared by Dr. K. Roy, University of Alberta. The sequences of each primer are as follows:

Set 1 spans the deleted region in the *ras*^{3} mutation:

Primer1 (5′ end) 5′ > ATAAAGAGGAAATGGATATTGCT < 3′
Primer2 (3′ end) 5′ > TTTGTGGATACAGATTCCTCTC < 3′

Set 2 spans the insertion sites in the *ras*^{1}, *ras*^{2} and *ras*^{4} mutations:

Primer3 (3′ end) 5′ > CGCATAAGAGTGTTGAGT < 3′
Primer4 (5′ end) 5′ > GGTAACCGAATCGGAGAT < 3′

2.5 RNA manipulation

2.5.1 RNA extraction and poly-A^+ RNA purification

Preparing and manipulating undegraded RNA samples is technically difficult and requires special laboratory precautions to prevent accidental RNase contamination (For a detailed discussion of the procedures, see Sambrook *et al.*, 1989). The following rules were followed:

a) Gloves were worn at all times;
b) Restricted "RNA only" working space, glassware, batches of plasticware, apparatus and chemicals were set apart from general working facilities;
c) Glassware was baked at 180°C overnight;
d) Water was treated by DEPC (diethyl pyrocarbonate);
e) Electrophoresis tanks were soaked with 3% H_2O_2 for at least an hour, then rinsed thoroughly with DEPC treated water,
f) RNA samples were kept in 0.5% SDS at -70°C.

Bulk RNA extraction was done using a modified single-step method of Chomczynski and Sacchi (1987). This method was originally a macro-method and was successfully scaled down for use with Eppendorf tubes, as follows:

a) Transfer ~50mg tissue samples from storage at -70°C to an ice cold "Micro Tissue Grinder" (Fisher) and immediately add 0.5ml RNA extraction buffer (room temperature). All the remain procedures were carried out at room temperature. Grind the samples thoroughly.
b) Transfer the homogenate to an Eppendorf tube and add 50µl 2M sodium acetate (pH4.0), 500µl water saturated phenol and 100µl chloroform:isoamyl alcohol mixture (49:1). For pupal and adult samples, an extra 50µl water was added. Vortex mix vigorously for 0.5-1 minute.
c) Centrifuge in a bench-top Eppendorf centrifuge at full speed for 15 minutes. Transfer the aqueous phase to a fresh Eppendorf tube. (If no distinct phases can be seen after centrifugation, another 100µl chloroform:isoamyl alcohol mixture is added and the mixture is centrifuged again after vortexing vigorously.) Add 500µl isopropanol (-20°C), mix well and place at -70°C for at least 1 hour.
d) Centrifuge in the Eppendorf centrifuge at full speed for 15 minutes. Resuspend the RNA pellet in 300µl extraction buffer, add 300µl isopropanol (-20°C), mix well and place at -70°C for at least 1 hour.
e) Centrifuge in the Eppendorf centrifuge at full speed for 15 minutes. Wash the RNA pellet with 1ml -20°C 70% ethanol twice. Vacuum dry. Add 100µl 0.5% SDS and keep at -70°C.

A preliminary estimate of the quantity and quality of purified RNA was obtained by crude agarose gel electrophoresis without glyoxal. Later, quality was confirmed mainly by considering the ribosomal bands on agarose gel electrophoresis of glyoxalated samples and RNA concentration was estimated by measuring the A_{260} derived from a 1µl sample. Ultimately, gel loading was determined (or readings of northern analyses adjusted) by measuring the *D. melanogaster* ribosomal protein 49 (RP49) transcript signal density after northern hybridization.

When necessary, poly-A^+ RNA was purified from total RNA by binding to oligo-(dT)-cellulose (Sambrook *et al.*, 1989).

2.5.2 RNA agarose gel electrophoresis, northern transfer

RNA agarose gel electrophoresis was done using the glyoxal/dimethyl sulfoxide (DMSO) method (Sambrook *et al.*, 1989).

Northern transfer employed BIOTRANS Nylon Membranes (ICN Biomedicals, Inc.). The capillary blot transfers were performed as described in the manufacturer's manual. Ethidium bromide was not added to the agarose gel when a transfer was contemplated. Rather, after electrophoresis, the RNA ladders (0.24-9.5kb RNA ladder, Gibco/BRL) were cut from the gel and stained for 30 minutes in 0.25M ammonium acetate, 5µg/ml Ethidium bromide, then destained 30-60 minutes in 0.25M ammonium acetate, in the dark. These stained size markers were then photographed under UV illumination alongside a ruler, for record purposes.

2.5.3 Ribo probes

Riboprobes were used in northern blot analysis of total RNA to improve the sensitivity of the method. It is effectively impossible to strip riboprobes from northern blots, limiting the useful life of each blot. However, the sensitivity gain associated with riboprobes allows analysis of total RNA, so that there is a compensating gain of convenience in avoidance of poly-A^+ selection of RNA samples.

The riboprobe was prepared using an "RNA Transcription Kit" (Stratagene). To prevent RNase contamination, all DNA templates were treated with Proteinase K, followed by phenol:chloroform extraction. Each transcription reaction was set up according to the manufacturer's manual. Ribonuclease inhibitor "RNase Block" (Stratagene) was also added to each reaction according to the manufacturer's suggestion. All probes were purified by passing through a Sephadex G-50 column (Sambrook *et al.*, 1989).

2.5.4 Hybridization and washing

When a riboprobe was used, the hybridization conditions were set up according to the suggestion provided in the transcription kit manual (50% formamide, 5X SSC, 1X PE and 150µg/ml denatured salmon sperm DNA, at 65°C, overnight).

When a DNA probe was used, hybridization was carried out at 65°C overnight with 50% formamide, 5X SSPE, 5X Denhardt's reagent, 1.0% (w/v) SDS and 100µg/ml nonhomologous DNA, a modification of the method in the BIOTRANS membrane manual (ICN Biomedicals, Inc.). The DNA probe was prepared using the same multiprime DNA labeling system (Amersham) used for Southern analysis (*Section 2.4.5.*). Generally, at least 10^6 cpm of probe/ml of hybridization solution were used.

No matter which kind of probe was used, washing was carried out under the same conditions: 2X SSC and 0.1% SDS at room temperature for 5 minutes twice followed by 0.1X SSC and 0.1% SDS at 65°C for 15 minutes three times.

2.6 Analysis of inosine monophosphate dehydrogenase activity

2.6.1 Protein extraction

Inosine monophosphate dehydrogenase (IMPDH) catalyzes the reaction from IMP to XMP. In this study, the IMPDH activity was analyzed by measuring the amount of ^{14}C-XMP that was derived from the substrate, ^{14}C-IMP ([8-^{14}C]Inosine 5′-monophosphate, Amersham), as a result of the activity of a crude Drosophila extract. The two radioactive nucleotides (^{14}C-IMP and ^{14}C-XMP) were separated by thin layer chromatography (TLC) and quantified using a PhosphorImager.

Crude protein samples were extracted from the tissues collected as described in *Section 2.3*. Frozen third instar larvae, pupae and adults were homogenized in 20mM Tris (pH7.5), 1mM EDTA and 0.5mM DTT, on ice. The homogenate was transferred to an Eppendorf tube and centrifuged at 4°C for 15 minutes at full speed, to remove debris. The crude protein extract was kept at -75°C to await analysis.

2.6.2 Protein quantification

The protein concentration in crude extracts was quantified using a Bio-Rad Protein Assay system (Bio-Rad Laboratories) based on the method described by Bradford (Bradford, 1976). This assay is a dye-binding assay in which Coomassie Brilliant Blue G-250 binds to protein. This binding causes a shift in the maximum absorption of the dye from 465 to 595nm and this shift is in response to various concentrations of protein. The increase in absorption at 595nm is monitored spectrophotometrically.

2.6.3 The assay reaction

The IMPDH activity assay was carried out in 100µl of reaction mixture (20mM Tris(pH8.0), 0.6mM EDTA, 20mM KCl, 1mM NAD^+ and 0.017mM ^{14}C-IMP) at 37°C in an Eppendorf tube. Initially, samples were taken from different time points, mainly in order to identify an approximately linear phase of the reaction. At each time point 10 to 30µl reaction mixture was transferred to an ice-cold Eppendorf tube containing an equal volume of 10% TCA, which stops the reaction and precipitates the proteins and other macromolecules. After mixing well on ice for 10 minutes and then centrifuging for 3 minutes at full speed, supernatant was transferred to a fresh tube and extracted with 500µl diethyl ether twice, to remove TCA.

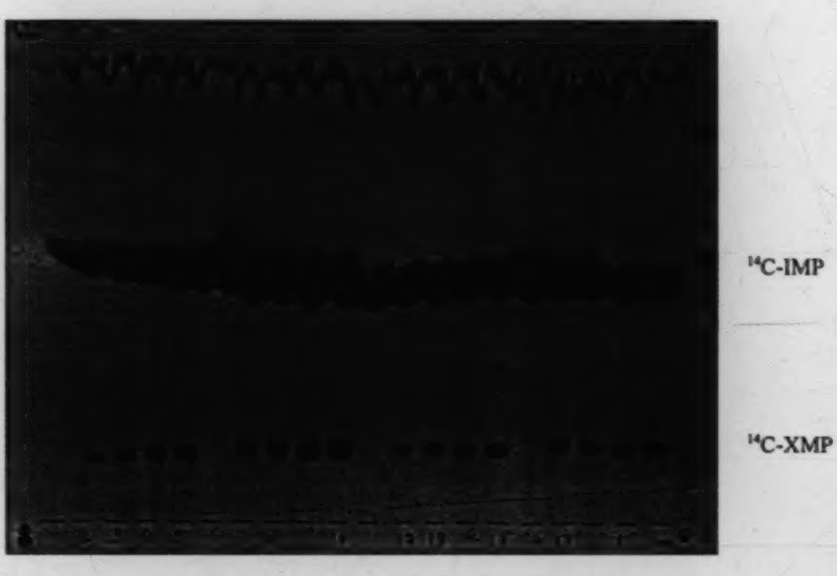
2.6.4 Chromatography

Polyethyleneimine (PEI) TLC plates were made with the help of Dr. K. Roy, University of Alberta, according to the method of Randerath and Randerath (1967). After drying at the room temperature, the coated plates were washed by ascending development with distilled water, a routine intended to remove impurities which might interfere with the chromatographic separation. The washed plates were kept at room temperature until used. The samples were loaded in 2-4μl aliquots and allowed to dry at room temperature. By repeated application, up to 30μl of the same sample could be loaded onto a single spot on the plate without undue spreading by capillary action. The loaded TLC plate was developed by two successive ascending runs to separate the nucleotides according to a method modified from Böhme and Schultz (1974). The first development was with distilled water. After drying in an air steam at room temperature the chromatograph was developed with 0.8M LiCl. In both runs, the solvents were allowed to run the entire length of the plate. For a 20cm TLC plate, each run takes about 1-2 hours. These running conditions were pre-tested and optimized for the maximal separation of the nucleotides IMP, XMP, AMP and GMP. Samples from the radioactive reaction were pre-run along with the cold IMP, XMP, AMP and GMP then checked under UV light. The ^{14}C-IMP and ^{14}C-XMP run at the same positions as cold IMP and XMP. Figure 2-4 shows an autoradiograph of a finished TLC plate. In the main body of the results, only the pattern of XMP production will be reported.

2.6.5 Phosphorimaging and computer analysis

The distribution of radioactivity on the TLC plates was ascertained by scanning in a PhosphorImager 445SI and analyzed using the software ImageQuant, both from Molecular Dynamics Inc. (Sunnyvale, CA).

Figure 2-4 Autoradiograph of a sample TLC plate



IMPDH activity analysis of four parallel reactions with the same Am^{+} protein sample which was extracted prior to the experiment. Samples were taken at different time points (shown in minutes). The reaction and TLC running conditions are described in the text. The finished TLC plate was exposed for 72 hours. The radioactivity on the top of the plate perhaps derived from the degraded ^{14}C labeled IMP.

The samples were taken at

1, 6, 11 and 16	0 minute of reaction
2, 7, 12 and 17	6 minutes of reaction
3, 8, 13 and 18	12 minutes of reaction
4, 9, 14 and 19	18 minutes of reaction
5, 10, 15 and 20	24 minutes of reaction

3 Results

3.1 Cloning of the *ras* gene

3.1.1 The Am$^+$ EMBL-4 library

An Am$^+$ genomic DNA library was made by ligating the 15-25kb *Sau* 3A digested Am$^+$ DNA fragments to the arms of the bacteriophage vector λ EMBL-4. In order to get inserts from regions with a whole range of *Sau* 3A restriction site densities, the 15-25kb Am$^+$ DNA fragments were collected from a series of samples ranging from DNA barely digested to DNA extensively digested with *Sau* 3A. This favours inclusion of most DNA sequences, from regions of the DNA that have very many *Sau* 3A sites to those that have few (Figure 2-2).

To test the whole system, experiments were carried out on a small scale prior to large scale ligation and plating. All the ligations were checked on mini-gels. The DNA was packaged with Packagene Extract System (Promega) before plating. The following preliminary tests were carried out:

a) To test the packaging system, pure EMBL-4 DNA was packaged and plated with either Q358 (high dilution) or NM539 (low dilution). Around $5\text{-}7\times10^6$ pfu/µg DNA were formed on the Q358 plates, but few plaques were found on NM539 plates.
b) The EMBL-4 DNA was digested with *Bam* HI, religated, packaged and plated with either Q358 or NM539. Similar results were found to those in the above experiment.
c) Prepared λ EMBL-4 arms (without the central fragment) were ligated on their own, packaged and plated with either Q358 or NM539. Less than 5×10^2 pfu/µg DNA were found on Q358 plates, and almost none were found on NM539 plates.
d) Different ratios of prepared EMBL-4 arms and *Sau* 3A digested Am$^+$ DNA fragments were ligated, packaged and plated with either Q358 or NM539. The optimum ratio was found at 1:1 between arms to inserts. This ratio gave the highest yield of about $1\text{-}3\times10^5$ pfu/µg DNA with both kinds of bacteria (Table 3-1).

The large scale ligations were performed at arms:inserts ratios of 1:0.5, 1:1 and 1:2, and plated only with NM539 cells. Combined, the number of plaques formed from the three ligations was approximately 400,000, which is the equivalent of about 5 to 10 genomes of *Drosophila melanogaster* (*Section 2.4.7.4*). These plaques were harvested and served as the amplified library used in this study.

3.1.2 Screening the library

The 8.4kb *Eco* RI fragment of LF1, which was from the P-element induced lethal mutant *ras*^{124}, contains wild-type *ras* DNA fragment with a P-element insert (Leonard, 1986). This 8.4kb fragment was used as the probe to screen the wild-type (Am$^+$) library. Amherst is a *ras*$^+$ M-strain; that is, it contains a normal *ras* gene and is P-element free.

Table 3-1 Tests of the genomic library cloning system

Table 3-1a* Packaging of naked EMBL-4 DNA

	Dilution						
Host cell	10^{0}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
Q358	—**	—	—	TMC**	TMC	54	7
NM539	5	1	0	0	—	—	—

* These tables demonstrate the efficiency of the cloning system. The undiluted concentration of λ phage EMBL-4 was about 1μg/100μl and referred as 10^{0} in the tables. All phage DNA was packaged prior to plating. Numbers in the tables are the averages of the recovered phages from two identically prepared plates.

** — = Not tested
TMC = Too Many to Count

Table 3-1b Religation of EMBL-4 vector

	Dilution						
Host cell	10^{0}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
Q358	—	—	—	—	TMC	68	5
NM539	4	0	0	—	—	—	—

Table 3-1c Estimation of vector contamination in vector arm preparations

	Dilution						
Host cell	10^{0}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
Q358	56	8	1	—	—	—	—
NM539	0	0	0	—	—	—	—

Table 3-1d Optimization of Arm:Inserts* ratio

	Dilution				
Arm:Inserts	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
4:1	TMC	TMC	250	19	6
2:1	TMC	TMC	286	25	4
1:1	TMC	TMC	329	29	5
0.5:1	TMC	TMC	303	22	5
0.25:1	TMC	TMC	290	29	4

* The concentration of the DNA insert was fixed at 0.2μg per reaction. Only the results generated from NM539 host cells are showing.

Aliquots of the amplified library described above were plated until a total of 400,000 plaques had been generated and screened. (Note, the similarity of the number of clones, 400,000, in the primary library and the sample screened is coincidental). Six positive clones were isolated.

3.1.3 Analysis of the wild type *ras* DNA

For convenience, the constructed chromosomal orientation and gene organization of the *ras* locus is illustrated in Figure 3-1.

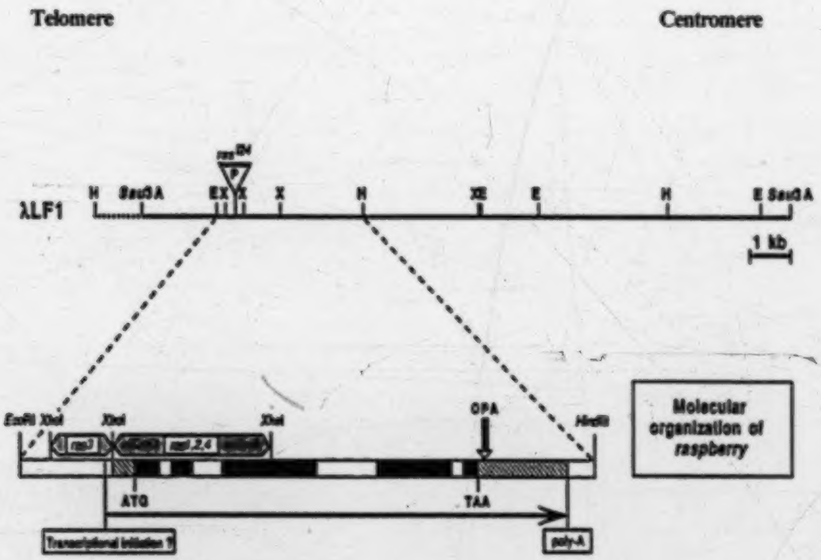
The six positive clones were analyzed by *Eco* RI, *Hin* dIII and *Xho* I digestion and, after Southern transfer, probed with different LF1 *Eco* RI fragments. Both restriction and Southern analysis showed that five of the six Am^+ phages were similar to LF1 (Figure 3-2). The major difference was that a 7.3kb *Eco* RI fragment replaced the 8.4kb fragment. This difference can be accounted for by assuming that a 1.1kb P-element is inserted in the ras^{D4} DNA. This observation agrees with Leonard's conclusion. All of the *Eco* RI Am^+ fragments in these phages were subcloned into both pUC19 and pBluescript plasmids.

In situ hybridization to salivary chromosomes showed that the 7.3kb *Eco* RI Am^+ fragment hybridized at the 9E region (*ras* locus) of the polytene chromosomes (Figure 3-3).

Restriction analysis revealed that there is one *Hin* dIII site in the 7.3kb ras^+ *Eco* RI fragment, which separates it into two fragments, 4.0kb and 3.3kb. These two fragments, along with other *Eco* RI fragments of the *ras* clones, were used to probe the genomic DNA of the three most studied *ras* mutants, ras^{D4}, ras^2 and *In(1)123* in the initial experiments. *In(1)123* is a *ras* lethal mutant which bears inversion breakpoints in or close to the *ras* locus and the centromere. The Southern analysis revealed that both ras^{D4} and ras^2 showed DNA alterations in the 4.0kb *Eco* RI-*Hin* dIII region (Figure 3-4). However, no DNA alteration was found in the whole 7.3kb *Eco* RI region in *In(1)123* (Figure 3-4). Leonard (1986) reported, on the evidence of *in situ* hybridizations, that the LF1 insert spanned the breakpoint of *In(1)123* and used this evidence as confirming the site of the P-element insertion as the *ras* locus. This observational paradox has not yet been resolved experimentally. It is known that rearrangements which, like *In(1)123*, place euchromatic genes close to centric heterochromatin, are commonly associated with breakpoints somewhat removed from the affected gene, a phenomenon known as position effect (Spofford, 1976), so the lack of an alteration in the immediate vicinity of the *ras* gene is not, in itself, paradoxical. Despite the *In(1)123* observation, the above data clearly show that this 7.3kb Am^+ *Eco* RI fragment, especially its 4.0kb *Eco* RI-*Hin* dIII sub-region, is associated with the *raspberry* gene.

The LF1 8.4kb ras^- *Eco* RI fragment contains an extra *Hin* dIII site in comparison to the 7.3kb ras^+ fragment. The extra *Hin* dIII site in LF1 is about 0.5kb from the *Eco* RI end in the ras^{D4} DNA (Leonard, 1986). Three *Xho* I sites are in the 4.0kb ras^+ *Eco* RI-*Hin* dIII fragment but four *Xho* I sites were found in the equivalent region of LF1 (Leonard, 1986). Both the extra *Hin* dIII and *Xho* I sites are presumably located in the inserted P-element of LF1 DNA. It is known that there is a *Hin* dIII site within the terminal repeats of the P-element (O'Hare and Rubin, 1983), implying that the 0.5kb *Eco* RI-*Hin* dIII fragment must span the insertion point of the P-element (Figure 3-5).

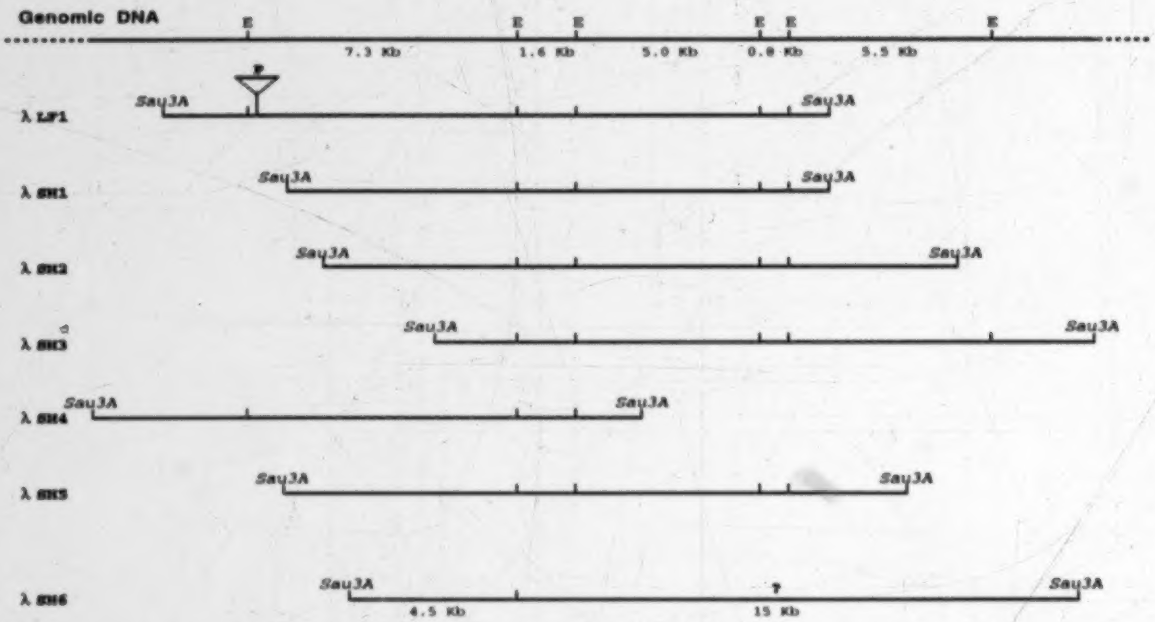
Figure 3-1 A genomic restriction map of the *ras* locus



The restriction map of the λ EMBL-4 clone LF1 derived from the mutant ras^{I24} is shown above (E, *Eco* RI; H, *Hin* dIII; X, *Xho* I). Only the *Sau*-3A sites at the ends of the insert DNA are shown. The *Hin* dIII site to the left is deduced from Southern analysis of wild-type genomic DNA. The 1.1kb P-element insert in ras^{I24} contains additional *Hin* dIII and *Xho* I sites.

The enlargement shows the genomic segment sequenced. Coding segments from a transcript (cDNA29) from the region are shown in black; the coding sequence is interrupted by four introns (unfilled). Untranslated segments of the cDNA are cross-hatched (Shaded boxes above the enlarged segment indicate the DNA segments affected by the four known spontaneous *ras* alleles). The situation regarding the four *ras* viable mutants as it was understood in 1994 is shown. The transcription unit probably commences upstream of the cDNA. (Adapted from Nash *et al.*, 1994)

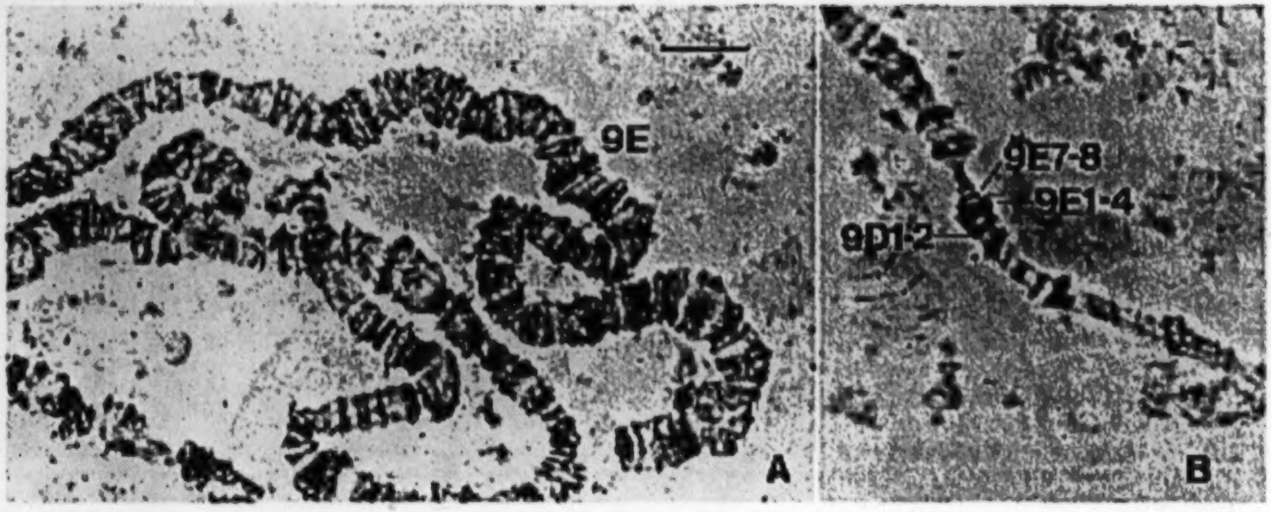
Figure 3-2 Restriction maps of λ LF1 insert DNA and the homologous λ EMBL-4 clones from the wild-type (Am+) genomic library



The Eco RI maps of the wild-type λ clones were derived from Southern analysis with the equivalent fragments in the λ LF1 phage as the probes. The 7.3kb wild-type Eco RI fragment which was extensively studied in this thesis was subcloned from SH4. The 15kb fragment of SH6 (with a question mark on the top) is not homologous to any of the fragments in λ LF1 and does not hybridize to the ras locus in the in situ hybridization experiment.

- E, *Eco* RI site, P, P-element

Figure 3-3 *In situ* hybridization of the Am^+ 7.3kb *Eco* RI fragment



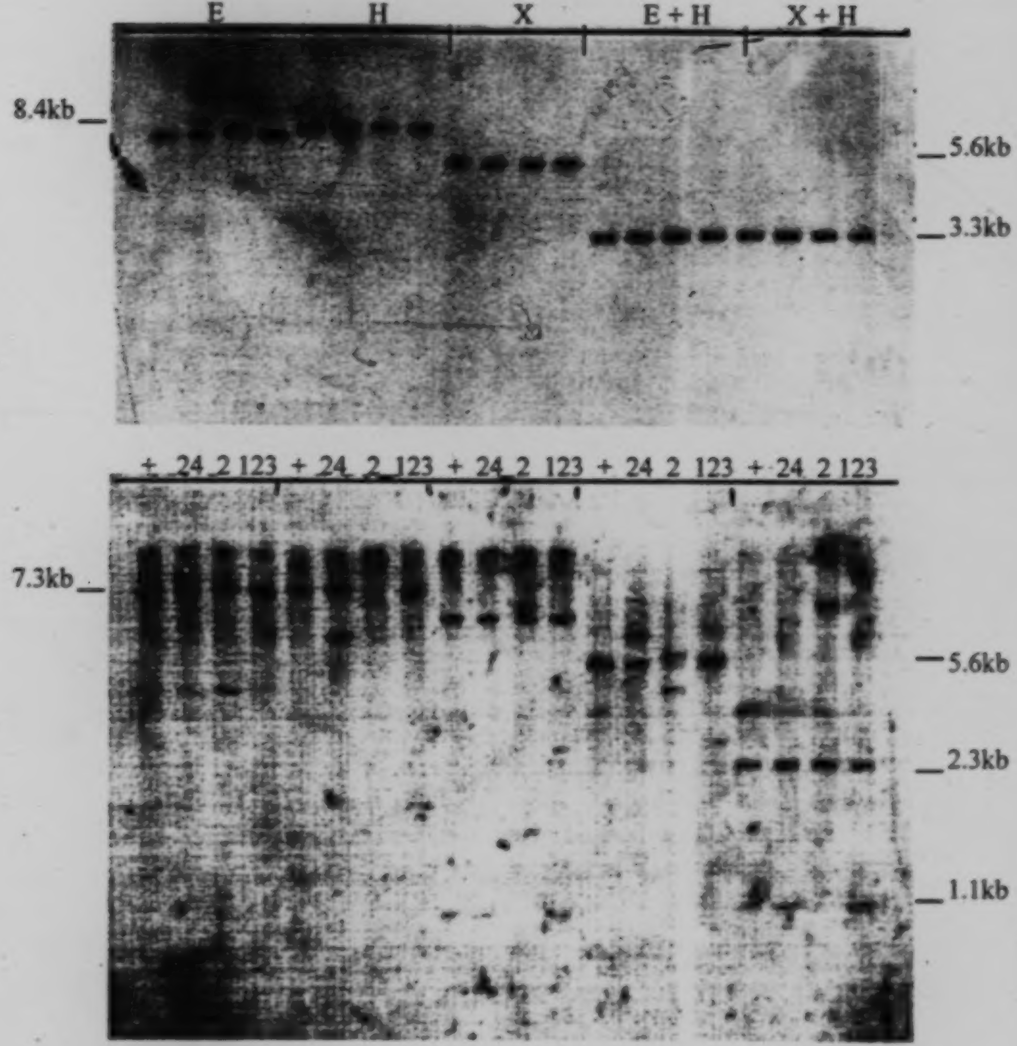
In situ hybridization of biotinylated the 4.0kb *Eco* RI-*Hin* dIII DNA fragment from the *ras* locus onto polytene chromosomes. (Scale bar = 10µm)

A. The signal at 9E is the only staining observed regularly on the X chromosome (as shown) or elsewhere.

B. Cytogenetic studies place raspberry in 9E3-4; however, even in well-stretched chromosomes staining does not seem restricted to 9E3-4 but extends to 9E1-2. It is clearly distinct from 9E7-8.

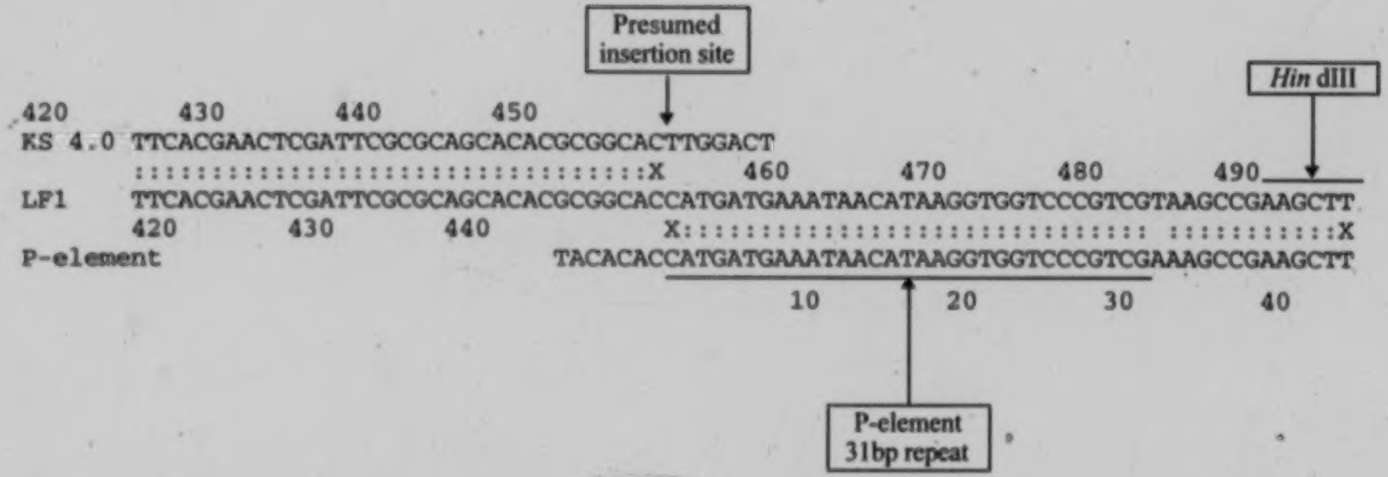
(From Nash *et al.*, 1994)

Figure 3-4 Southern analysis of Am$^+$ 3.3 and 4.0kb *Eco* RI fragments



Parallel Southern analyses of Am$^+$ (+), *ras^{124}/FM6* (24), *ras^2* (2) and *In(l)123/FM6* (123) genomic DNA which were digested by *Eco* RI (E), *Hin* dIII (H), *Xho* I (X), *Eco* RI plus *Hin* dIII (E+H) and *Xho* I plus *Hin* dIII (X+H). The 3.3kb and 4.0kb sub-fragments of the 7.3kb wild-type *ras* DNA were used as probes. Note there is no visible alterations in *In(l)123/FM6*. Top: Probed by the 3.3 *Hin* dIII-*Eco* RI fragment. The 7.3kb fragment in *ras^{124}* and *ras^2* was altered. No alteration was found in the 3.3kb *Hin* dIII-*Eco* RI fragment, the 8.4kb *Hin* dIII fragment and the 5.6kb *Xho* I, indicating the alteration is near the *Eco* RI site of the 4.0kb *Eco* RI-*Hin* dIII fragment. Bottom: Probed by the 4.0kb *Eco* RI-*Hin* dIII fragment. The pattern of the 7.3 *Eco* RI fragment is the same as in A. The 1.1kb *Xho* I fragment in *ras^2* was altered, indicating the *ras^2* mutation probably occurred in this region.

Figure 3-5 The sequence at the 5' end of the P-element insertion in λ LF1



```
420       430       440       450
KS 4.0   TTCACGAACTCGATTCGCGCAGCACACGCGGCACTTGGACT
         ::::::::::::::::::::::::::::::::::X     460       470       480       490___
LF1      TTCACGAACTCGATTCGCGCAGCACACGCGGCACCATGATGAAATAACATAAGGTGGTCCCGTCGTAAGCCGAAGCTT
         420       430       440           X::::::::::::::::::::::::::::::: :::::::::::X
P-element                            TACACACCATGATGAAATAACATAAGGTGGTCCCGTCGAAAGCCGAAGCTT
                                              10        20        30        40
```

The diagram shows the sequences spanning the P-element insertion in *ras*^{l24} (NJ5). The wild-type sequences are generated from the 4.0kb *Eco* RI-*Hin* dIII fragment. The *ras*^{l24} sequences are generated from the 497bp *Eco* RI-*Hin dIII* fragment of the 8.4kb *Eco* RI fragment of λ LF1. The P-element sequence (pπ25.1, which was used to induce *rasl*^{l24}) is from O'Hare and Rubin (1983).

Comparison of the sequence of λ SH4 4.0kb *Eco* RI-*Hin* dIII fragment and λ LF1 0.5kb *Eco* RI-*Hin* dIII fragment shows that they share the first 453bp sequences.

Comparison of the sequence of P-element and the λ LF1 0.5kb *Eco* RI-*Hin* dIII fragment shows that they share 44bp sequences, including the 31bp perfect repeat sequence of the P-element.

The P-element sequence in LF1 is adjacent to the Drosophila sequence, where the insertion presumably occurred.

The 0.5kb *Eco* RI-*Hin* dIII fragment was gel purified from the 8.4kb pUC19 plasmid and subcloned into pBluescript. Sequencing from both ends revealed that the P-element was inserted 453bp from the *Eco* RI site, about 150bp upstream from a potential transcription start site (see below, *Section 3.2*). There is a 44bp sequence of P-element in the 0.5kb fragment. The first 31bp of the *ras*^{D4} insert is identical to the "31bp perfect inverse terminal repeats" sequence of the P-element (O'Hare and Rubin, 1983). There is only one base pair mismatch (A/T) between the 44bp sequence of *ras*^{D4} insert and the first 44bp of the P-element sequence in pπ25.1, which was used in the original P-M mutagenesis experiment (Figure 3-5).

In situ hybridization of the sixth Am^{+} phage showed that it contained two noncontiguous DNA fragments (data not shown) and was not studied further.

3.1.4 Isolation of *ras* cDNA clones

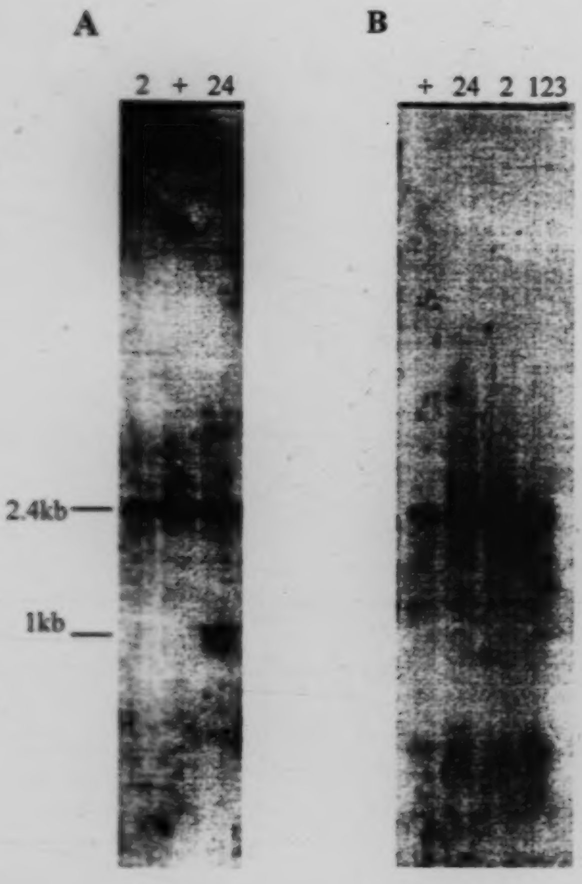
The 7.3kb Am^{+} *Eco* RI fragment is associated with the *ras* gene. To determine if there are any transcripts from this 7.3kb region, the 7.3kb fragment was used to probe female poly-A^{+} RNA of wild-type and the *ras*^{D4}, *ras*^{2} and *In(l)123* mutants. A northern analysis showed that the 7.3kb *Eco* RI fragment predominantly hybridized to a 2.4kb band (Figure 3-6). This indicates sequences within the 7.3kb *Eco* RI fragment are transcribed.

Using the same probe to screen an embryonic Oregon-R λgt10 cDNA library (E6-8) obtained form Dr. T. Kornberg (Poole *et al.*, 1985), seven positive clones were isolated. One of them, cDNA4 contains an internal *Eco* RI site. To analyze the relationship among them, they were cut by the four base-pair cutters *Sau* 3A and *Hae* III, which generate short restriction fragments. All seven clones share some common *Sau* 3A bands, even though they are clearly different from each other (Figure 3-7). Southern analysis showed that these cDNA clones, except cDNA1 and the small Eco RI fragment of cDNA4, hybridized to the 4.0Kb *Eco* RI-*Hin* dIII fragment, but not the 3.3kb *Hin* dIII-*Eco* RI fragment, indicating that they are transcribed from the 4.0kb *Eco* RI-*Hind* III region (Figure 3-7).

All the cDNA inserts were re-cloned into pBluescript plasmid. The sequence of both strands of cDNA2 (2.4kb, SK29) revealed that it contains an open reading frame (ORF) encoding an IMPDH-like protein. cDNA10 (2.3kb, SK83) was only sequenced from one strand. SK83 contains the same ORF, although it is ~200bp shorter at the 5' end and contains some sequence differences from SK29 (The sequence data will be discussed fully in the next section). Partial sequence of cDNA8 (2.3kb, SK16) revealed that it contains the same 5' end as SK29, suggesting it is also an IMPDH-like transcript.

Three cDNAs, cDNA3 (1.5kb, SK49), cDNA4 (2.9kb, SK5 and SK7) and cDNA7 (0.9kb, SK55) were partially sequenced only from the universal primer of the SK(+) phagemid. The partial sequences obtained from these three cDNAs were complementary to the sequence obtained from the 3' end of SK29. No further experiment has been done to characterize these mutants. They may be artificial effects of the cDNA library construction, or may be generated from the mRNAs transcribed from the opposite direction of SK29. When using a riboprobe synthesized from the first 800bp of KS29, only a predominant 2.4kb transcript was observed in all Northern analyses. Slee and Bownes (1995) reported that two IMPDH

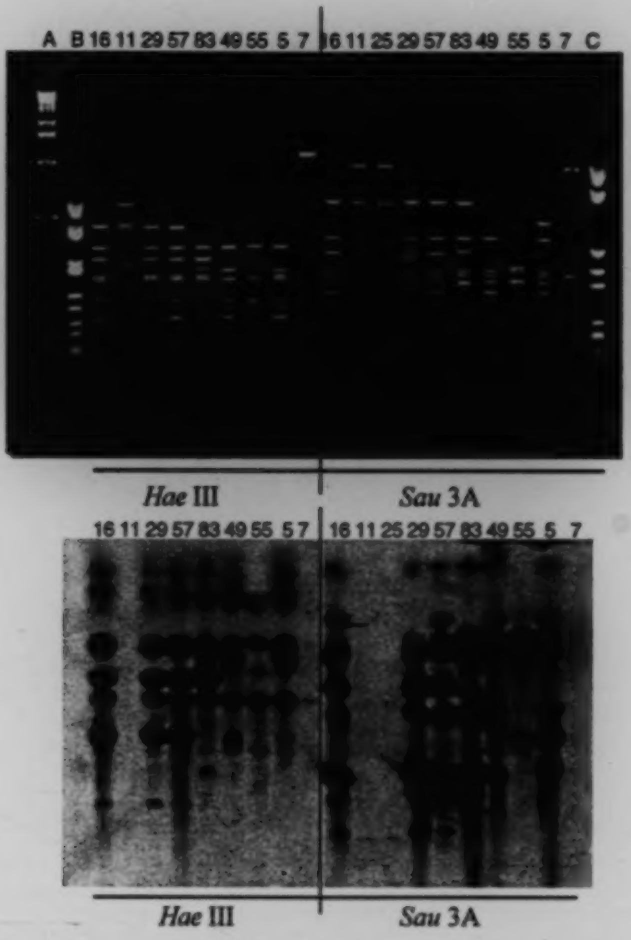
Figure 3-6 Poly-A$^+$ RNA probed by *ras* DNA fragments



A. Female poly-A$^+$ RNA from *ras*^{2} (2), Am$^+$ (+) and *ras*^{l24}*/FM6* (24) probed by λ LF1 8.4kb fragment. A highly labeled 2.4kb band is present in all three samples. There is an additional small transcript in *ras*^{l24}*/FM6*.

B. Female poly-A$^+$ RNA from Am$^+$ (+), *ras*^{l24}*/FM6* (24), *ras*^{2} (2) and *In(1)123/FM6* (123) probed with a λ SH4 7.3kb fragment. The same 2.4kb transcript is present in all four samples. The small transcript found in *ras*^{l24}*/FM6* (in A) is absent when probed with wild-type DNA, indicating it to be derived from P-element transcription. The normal signal in the strains may be derived exclusively from the FM6 chromosome.

Figure 3-7 Analysis of *ras* putative cDNAs digested by restriction enzymes with four base-pair target sites



Top: The λ gt10 cDNA clones originally selected with the 7.3kb *Eco* RI fragment from λ SH4 were re-cloned into pBluescript. The subclones were named as pSKXX based on the arbitrary number assigned to them during subcloning. The pBluescript subclones were amplified and the cDNA inserts were isolated. In order to perform a rapid sequence comparison between the various cDNAs, the purified cDNA inserts were digested by enzymes with four base-pair target sites (either *Hae* III or *Sau* 3A) and run in 2% agarose gel at 200V for 3.5 hours. Subclone 11 and 25 were cloned from the same cDNA insert. Subclone 5 and 7 were generated from two *Eco* RI fragments of the same λ cDNA insert.

In the figures,

A = 1kb DNA ladder (BRL); B = pBluescript cut by *Hae* III C; = pBluescript cut by *Sau* 3A
16 = pSK16; 11 = pSK11; 29 = pSK29; 57 = pSK57; 83 = pSK83
49 = pSK49; 55 = pSK55; 5 = pSK5; 7 = pSK7; 25 = pSK25

Bottom: The DNA in above gel was transferred to the filter and probed by the 4.0kb *Eco* RI-*Hin* dIII and 3.3kb *Hin* dIII-*Eco* RI fragments of λ SH4 respectively.

mRNAs (2.5kb and 1.9kb) were highly expressed in ovaries and another 0.9kb transcript was specifically expressed in testes. However, since they used a DNA probe which could have hybridized to transcripts of both orientations, their signals could have been synthesized from either direction.

No homology with the other cDNAs was found in cDNA1 (2.8kb, SK11). Information describing these cDNAs is listed in Table 3-2.

3.2 Sequence analysis of the *ras* gene

A detailed description of this work is to be found in Nash *et al.* (1994), and significant segments of the text in the following sections (*Section 3.2, Section 3.3.1*) are derived from that publication.

3.2.1 Exon/intron structure

Southern analysis (Nash *et al.*, 1994) revealed that the 4.0kb *Eco* RI-*Hin* dIII fragment that contained the P-element insert in λ LF1 was altered in all five P-element induced *ras* lethal mutants (Figure 3-12) and in all four spontaneous *ras* viable mutants (Figure 3-13). No other alteration was observed in the remainder of the cloned DNA from the *ras* region. This suggests that the 4.0kb *Eco* RI-*Hin* dIII region of the *ras* locus is the location of all or most of the *ras*$^+$ gene. To investigate the *ras*$^+$ gene organization, the 4.0kb genomic fragment and the longest cDNA, cDNA29, were sequenced from both strands. Both sequences are shown in Figure 3-8. It is also available under Genbank accession #L14847. In Figure 3-8, the A in the putative translation start codon ATG is allocated the number one (+1) in the genomic sequence.

The 3976bp genomic sequence contains sequences equivalent to that from the sense-strand of the 2361bp cDNA29 transcript, interrupted by four introns. There is a single base difference internal to the matched sequences (G *vs.* A at position 2662), in an apparently untranslated region.

Of the four introns, only the fourth (2208-2278), at 71bp, is within the size range (65-75bp) commonly found for *D. melanogaster* introns (Mount 1982). Intron 3 (1254-1666) is relatively large at 413bp and contains a possible transcription termination signal, TTTTCAAAA (1395-1403) (Henikoff and Eghtedarzadeh, 1987), and a polyadenylation signal, AATAAA (1525-1530) (Proudfoot and Brownlee, 1976), which are, however, unusually widely separated. Whether they are used as alternative transcription termination signals is unknown. All four introns fall within the open reading frame. The introns have conventional GT and AG internal boundaries.

3.2.2 Transcription and translation start and termination sites

The 5′ regulatory sequences of the *ras* gene are shown in Figure 3-9. Given the limited accuracy of estimates of mRNA sizes from northern transfers, cDNA 29 is potentially full-length. However, the sequence ATCTTCT (starting at -197) is an approximation of the consensus transcriptional start site found in many insect genes, ATCA(G/T)T(C/T) (Hultmark *et al.*, 1986). It is likely, therefore, that cDNA 29, which starts at -140, is missing 57bp at its 5′-end, relative to the corresponding mRNA. No

Table 3-2 The relationship of the cDNAs

Phage number	Insert size (kb)	Plasmid number	5' end sequence relative to SK(+) 29
cDNA 1#	2.8	SK(+) 11, 25	no homology
cDNA 2#	2.4	SK(+) 29	all
cDNA 3#	1.5	SK(+) 49	3' end
cDNA 4#	1.8	SK(+) 5	middle
	1.1	SK(+) 7	no homology
cDNA 7#	0.9	SK(+) 55	3' end
cDNA 8#	2.3	SK(+) 16, 57	5' end
cDNA 10#	2.3	SK (+) 83	5' end

Figure 3-8 Comparison of the nucleotide sequence of the 4.0kb *Eco* RI-*Hin* dIII fragment from λ SH4 and cDNA pSK29

The 4.0kb *Eco* RI-*Hin* dIII *(ras)* DNA sequence

```
 Eco RI
gaattcggtg tgcagccaca cagatgcgat atttatcgat agtctcggtg   -752

agctgtgcac gaagccagcc atgaaacagt taactgtgac acttctatcg   -702

ttactatcga tagctggcat tgtaagcaat agtatcgcag tgcagcaata   -652

gctttaagcg gaaagaagct aattcgctgc catattttaa agattaaata   -602

aagaggaaat ggatattgct cgagattcag gctattcagg ctaatatgta   -552

atgattattg tttaatatgc ttaataaatt gcatagtcga tattgaaagc   -502

gccttagtgt ggctgccatg gaaagcgcga ttaaggccaa agcgtatcgt   -452

ggtaatattt agtgggtttc ccggtggtcc agcgagtata tttaagcggc   -402

cgcacggcgg ccacaccact tcacgaactc gattcgcgca gcacacgcgg   -352

cacttggact ccagcggact ggtaagatgg cccaaccatg tgctcgctgg   -302

cttgccagcc aatcgggccg attcccattg cgaatcaccc tccgatcgat   -252

gcaattaacc ttttccattt gatcccaatt gtctggctgc gccgtgtgcc   -202

gtgcatcttc ttcactgccg catttgtgta tgtgccccct gctaattgcg   -152
potential transcription start site
tctcaaaagc tcgagaggaa tctgtatcca caaaaagaaa tctgcaacag   -102
    cDNA29  cgagaggaa tctgtatcca caaaaagaaa tctgcaacag

gagaagatcc caaaacaggc ggtttttatc tctatacggc ttgttattgt    -52
gagaagatcc caaaacaggc ggtttttatc tctatacggc ttgttattgt

cagctccccg ttgtttgttg actgtttgtt tgggttgaac gcacgcggaa     -2
cagctccccg ttgtttgttg actgtttgtt tgggttgaac gcacgcggaa

aATGGAGAGC ACGACAAAGG TGAAGGTGAA CGGCTTTGTG GAATCGACGT     49
aATGGAGAGC ACGACAAAGG TGAAGGTGAA CGGCTTTGTG GAATCGACGT

CGTCTTCAGC GGCGCCGGCA ATCCAGACAA AGAGTACCAC CGGATTCGAT     99
CGTCTTCAGC GGCGCCGGCA ATCCAGACAA AGAGTACCAC CGGATTCGAT

GCCGAGCTGC AGGATGGGCT GAGTTGTAAG GAACTGTTCC AGAACGGTGA    149
GCCGAGCTGC AGGATGGGCT GAGTTGTAAG GAACTGTTCC AGAACGGTGA

GGGACTCACC TACAAgtgag tagtgatcaa gcctctgcaa aagtcgttat    199
GGGACTCACC TACAA

ccaaatctca tgtcattgcc cccatccgaa attcccgaca atcccgaaaa    249
```

```
aaaacagCGA CTTTCTCATA CTGCCCGGCT ACATAGACTT CACCGCCGAG    299
       CGA CTTTCTCATA CTGCCCGGCT ACATAGACTT CACCGCCGAG

GAGGTCGATC TCAGTTCGCC ACTGACCAAG TCGCTGACAT TGCGAGCACC    349
GAGGTCGATC TCAGTTCGCC ACTGACCAAG TCGCTGACAT TGCGAGCACC

GCTGGTTAGT TCGCCCATGG ACACGGTAAC CGAATCGGAG ATGGCCATCG    399
GCTGGTTAGT TCGCCCATGG ACACGGTAAC CGAATCGGAG ATGGCCATCG

CCATGGCGgt aagtatcgga ttttaatgaa tccctagaca tacatatata    449
CCATGGCG

cacatatata tacggcccag aaccgctggc gggttcgcct gagctgtgct    499

agcctgtggt tgtaactcaa cactcttatg cgcctctcta accgccaatg    549

tttctatcca ctgctcacac accatccaaa tccattcgaa tcccgcgcgc    599

agCTGTGTGG TGGCATTGGC ATCATCCATC ACAACTGCAC GCCGGAATAC    649
  CTGTGTGG TGGCATTGGC ATCATCCATC ACAACTGCAC GCCGGAATAC

CAGGCGTTGG AGGTGCACAA GGTTAAGAAG TACAAGCACG GCTTCATGCG    699
CAGGCGTTGG AGGTGCACAA GGTTAAGAAG TACAAGCACG GCTTCATGCG

CGACCCCTCG GTGATGTCGC CCACGAATAC GGTGGGGGAT GTGTTGGAGG    749
CGACCCCTCG GTGATGTCGC CCACGAATAC GGTGGGGGAT GTGTTGGAGG

CGCGGCGGAA GAACGGATTC ACCGGCTATC CGGTCACCGA GAACGGCAAA    799
CGCGGCGGAA GAACGGATTC ACCGGCTATC CGGTCACCGA GAACGGCAAA

CTTGGCGGCA AGCTGCTGGG CATGGTCACG TCGCGAGACA TTGACTTCCG    849
CTTGGCGGCA AGCTGCTGGG CATGGTCACG TCGCGAGACA TTGACTTCCG

CGAGAACCAG CCGGAGGTCC TGCTGGCCGA CATCATGACC ACCGAACTGG    899
CGAGAACCAG CCGGAGGTCC TGCTGGCCGA CATCATGACC ACCGAACTGG

TCACCGCTCC CAATGGCATC AATCTGCCCA CGGCAAACGC TATTCTCGAG    949
TCACCGCTCC CAATGGCATC AATCTGCCCA CGGCAAACGC TATTCTCGAG

AAGAGCAAGA AGGGCAAACT GCCGATTGTG AATCAGGCCG GCGAACTGGT    999
AAGAGCAAGA AGGGCAAACT GCCGATTGTG AATCAGGCCG GCGAACTGGT

GGCCATGATT GCCCGCACGG ATTTGAAGAA GGCCCGCTCC TACCCGAATG  1,049
GGCCATGATT GCCCGCACGG ATTTGAAGAA GGCCCGCTCC TACCCGAATG

CCTCCAAGGA CTCCAATAAG CAGCTCCTCG TCGGCGCTGC CATCGGAACG  1,099
CCTCCAAGGA CTCCAATAAG CAGCTCCTCG TCGGCGCTGC CATCGGAACG

CGATCGGAGG ACAAGGCGCG CCTGGCTCTG CTGGTGGCCA ACGGCGTGGA  1,149
CGATCGGAGG ACAAGGCGCG CCTGGCTCTG CTGGTGGCCA ACGGCGTGGA

CGTTATCATC CTTGATTCAT CGCAGGGCAA CTCCGTCTAC CAGGTGGAGA  1,199
CGTTATCATC CTTGATTCAT CGCAGGGCAA CTCCGTCTAC CAGGTGGAGA

TGATCAAGTA CATTAAGGAG ACCTATCCCG AACTACAGGT CATTGGTGGC  1,249
TGATCAAGTA CATTAAGGAG ACCTATCCCG AACTACAGGT CATTGGTGGC

AATGgtaagt accagaacac tacgtaaaag ggtatttcaa aaatgaagac  1,299
AATG

acatatagta tacacaggtc attcgcttgt ccgtatgaat ttgagatgtt  1,349

tttaagttca atgctcttaa agtatcacat tttgatatac tcacgttttc  1,399
                                          termination?
```

```
aaaactattt ttaaccatta tctatggcta attacgtcgc gatgaagtac 1,449

ttattattat agcaatcatg atttggagtc aaacatattg ttatctttag 1,499

tatcacgaaa tgacaattgg gttgaaataa agatagggtg caagtttta  1,549
                        polyadenylation?
aacagatttc cagttaattg acataattat tatattgatt ataatcaagt 1,599

gattttaatt tattgttact tgcaaatttt ccactaattc aaaatattta 1,649

ttttattcac tgaacagTGG TAACACGTGC CCAGGCCAAA AATCTCATTG 1,699
                  TGG TAACACGTGC CCAGGCCAAA AATCTCATTG

ATGCCGGCGT CGATGGACTG CGTGTGGGCA TGGGTTCCGG TTCCATCTGC 1,749
ATGCCGGCGT CGATGGACTG CGTGTGGGCA TGGGTTCCGG TTCCATCTGC

ATCACCCAGG AGGTGATGGC CTGCGGATGT CCTCAGGCCA CTGCTGTCTA 1,799
ATCACCCAGG AGGTGATGGC CTGCGGATGT CCTCAGGCCA CTGCTGTCTA

CCAGGTGTCC ACGTACGCTC GCCAATTCGG AGTGCCAGTG ATCGCGGACG 1,849
CCAGGTGTCC ACGTACGCTC GCCAATTCGG AGTGCCAGTG ATCGCGGACG

GAGGCATCCA GTCCATTGGA CACATTGTGA AGGCTATTGC ATTGGGAGCC 1,899
GAGGCATCCA GTCCATTGGA CACATTGTGA AGGCTATTGC ATTGGGAGCC

AGCGCTGTAA TGATGGGCTC CCTGCTGGCG GGCACATCGG AGGCACCGGG 1,949
AGCGCTGTAA TGATGGGCTC CCTGCTGGCG GGCACATCGG AGGCACCGGG

CGAGTACTTC TTCTCCGACG GAGTGCGCCT TAAGAAGTAC CGCGGCATGG 1,999
CGAGTACTTC TTCTCCGACG GAGTGCGCCT TAAGAAGTAC CGCGGCATGG

GCTCCCTGGA GGCCATGGAG CGCGGCGATG CCAAGGGCGC TGCCATGTCG 2,049
GCTCCCTGGA GGCCATGGAG CGCGGCGATG CCAAGGGCGC TGCCATGTCG

CGCTACTACC ACAACGAGAT GGATAAAATG AAGGTGGCAC AGGGCGTCAG 2,099
CGCTACTACC ACAACGAGAT GGATAAAATG AAGGTGGCAC AGGGCGTCAG

TGGCAGCATC GTGGACAAGG GCAGTGTGCT GCGCTATCTA CCCTATTTGG 2,149
TGGCAGCATC GTGGACAAGG GCAGTGTGCT GCGCTATCTA CCCTATTTGG

AGTGCGGACT GCAGCACAGT TGTCAGGACA TCGGTGCTAA CTCCATTAAT 2,199
AGTGCGGACT GCAGCACAGT TGTCAGGACA TCGGTGCTAA CTCCATTAAT
                                          polyadenylation?
AAATTGAGgt aggctaacat ggacctaatt ccaaagaaca tgaaactaat 2,249
AAATTGAG

ctttatctcc gcatgttaaa tttccgcagG GATATGATCT ATAATGGCCA 2,299
                               G GATATGATCT ATAATGGCCA

ACTGCGCTTC ATGAAGCGCA CCCATTCCGC TCAACTGGAA GGCAATGTCC 2,349
ACTGCGCTTC ATGAAGCGCA CCCATTCCGC TCAACTGGAA GGCAATGTCC

ATGGCCTCTT CAGCTACGAG AAGCGTCTCT TCTAAcatac gacgatggct 2,399
ATGGCCTCTT CAGCTACGAG AAGCGTCTCT TCTAAcatac gacgatggct
                              stop translation
gcgacggaaa cgggcagtgg aggagtgcct gcccctgtcc agcagcagca 2,449
gcgacggaaa cgggcagtgg aggagtgcct gcccctgtcc agcagcagca

gcagcagcag cagcagcaac accatcagca gcaacagcat cgccaccgat 2,499
gcagcagcag cagcagcaac accatcagca gcaacagcat cgccaccgat
  imperfect trinucleotide repeat
ctgcagcagg cccactgctc gactacaccc gcctattcga tcgcaagtga 2,549
ctgcagcagg cccactgctc gactacaccc gcctattcga tcgcaagtga
```

```
gagaagaaaa gagaatcgag cgctcgaata gcgcatcctt ttcccctgga  2,599
gagaagaaaa gagaatcgag cgctcgaata gcgcatcctt ttcccctgga

tttttccaa tgttaggtta agtttaatcg gaatcgctga cgttaaagat  2,649
tttttccaa tgttaggtta agtttaatcg gaatcgctga cgttaaagat

tacgattcca atgtagattc cccttttga tggcggtgga ggttacctcg  2,699
tacgattcca atatagattc cccttttga tggcggtgga ggttacctcg
             ^
aatcgcccga atacatgtgt atacttaatc catctgtctg gtagacagaa  2,749
aatcgcccga atacatgtgt atacttaatc catctgtctg gtagacagaa

tcctaattat aagtgtttgc aggttacttg atctgtatag aaagtatata  2,799
tcctaattat aagtgtttgc aggttacttg atctgtatag aaagtatata

atcttcaact attcgctata tatatctaaa gcgcaacgcg ttggttggcc  2,849
atcttcaact attcgctata tatatctaaa gcgcaacgcg ttggttggcc

gcttacttat tttgcacgct atatgtatat agatgttaat ctctatttgt  2,899
gcttacttat tttgcacgct atatgtatat agatgttaat ctctatttgt

atagtgttgg atgtttacta ttataccgaa gctagcttca agttttaaaa  2,949
atagtgttgg atgtttacta ttataccgaa gctagcttca agttttaaaa
                              transcription termination signal

ctctgccaaa cttattaaat gtgaactgaa actaactgaa aagtatctcc  2,999
ctctgccaaa cttattaaat gtgaactgaa actaactgaaaaaaaaaaaaaaa
        polyadenylation signal
aagcggcttt taatttaata cctcggaggt caaccttatc tgcactctat  3,049

gaaaggggta tagaccttat cagctcgccc aataatgtaa tatgttttcg  3,099

gtcgttgtta tcaattgtgg acactgggat tctggctgtt ggccgattta  3,149

gtatgcgcaa gtttgttttt cgccaaaagc tt                     3.181
                            Hin dIII
```

The genomic DNA sequence is numbered from the A in the ATG codon at the presumed translation start. The cDNA sequence is aligned to the equivalent genomic DNA sequence. Within the exonic sequences, there is a single mismatch, at position 2662. The open reading frame (coding region) is capitalized. Registered as GenBank accession No. L14847.

(Adapted from Nash *et al.*, 1994)

Figure 3-9 Features in the "upstream" sequences of *ras* gene

5′ upstream sequences of *ras* gene

```
gaattcggtg tgcagccaca cagatgcgat atttatcgat agtctcggtg   -752
 Eco RI
agctgtgcac gaagccagcc atgaaacagt taactgtgac acttctatcg   -702

ttactatcga tagctggcat tgtaagcaat agtatcgcag tgcagcaata   -652

gctttaagcg gaaagaagct aattcgctgc catattttaa agattaaata   -602

aagaggaaat ggatattgct cgagattcag gctattcagg ctaatatgta   -552

atgattattg tttaatatgc ttaataaatt gcatagtcga tattgaaagc   -502

gccttagtgt ggctgccatg gaaagcgcga ttaaggccaa agcgtatcgt   -452

ggtaatattt agtgggtttc ccggtggtcc agcgagtata tttaagcggc   -402

cgcacggcgg ccacaccact tcacgaactc gattcgcgca gcacacgcgg   -352

cacttggact ccagcggact ggtaagatgg cccaaccatg tgctcgctgg   -302
  ^ P element insertion site in LF1
cttgccagcc aatcgggccg attcccattg cgaatcaccc tccgatcgat   -252

gcaattaacc ttttccattt gatcccaatt gtctggctgc gccgtgtgcc   -202
   ▼▼
gtgcatcttc ttcactgccg catttgtgta tgtgccccct gctaattgcg   -152
potential transcription start site

tctcaaaagc tcgagaggaa tctgtatcca caaaaagaaa tctgcaacag   -102
   cDNA29   cgagaggaa tctgtatcca caaaaagaaa tctgcaacag

gagaagatcc caaaacaggc ggtttttatc tctatacggc ttgttattgt   -52
gagaagatcc caaaacaggc ggtttttatc tctatacggc ttgttattgt

cagctccccg ttgtttgttg actgtttgtt tgggttgaac gcacgcggaa   -2
cagctccccg ttgtttgttg actgtttgtt tgggttgaac gcacgcggaa

aATGGAGAGC ACGACAAAGG TGAAGGTGAA CGGCTTTGTG GAATCGACGT   49
aATGGAGAGC ACGACAAAGG TGAAGGTGAA CGGCTTTGTG GAATCGACGT
start translation
```

The 5′ upstream sequence is taken from Figure 3-8 up to the translation start. The GC rich region is indicated as the following: The *Italic* letters represent 63.6% GC and the underlined letters represent 71.4% GC content. The postulated initiator is labeled. The two arrows indicate the consensus -1C and +1A (bold) in a typical initiator. The P-element insertion site in ras^{l24} is also included.

TATA box was found in the upstream 5′ region. The CA at the -1 and +1 positions of the putative transcriptional start site seems to be the conserved CA dinucleotide of an initiator. There is a 56bp GC rich region upstream of this putative initiator site (-351 to -406), with 71.4% GC compared to the 48.3% GC in the entire 4.0kb fragment. This 56bp is part of an 151bp moderately GC rich stretch with a 63.6% GC content (-282 to -432).

The cDNA includes an extensive open reading frame, which is presumed to be translated from the first in-frame ATG codon. Immediately upstream is the sequence GAAA, which can be compared with the consensus in this position, (C/A)AA(C/A) (Cavener, 1987). The open reading frame ends at a TAA triplet at 2381-3 (genomic numbering). The cDNA contains a poly-A tail approximately 60 residues long. Just upstream of the poly-A sequence, is a polyadenylation signal sequence, ATTAAA (2963-8) (Proudfoot and Brownlee, 1976) and a transcription termination signal, TTTTAAAA (2942-9) (Henikoff and Eghtedarzadeh, 1987).

3.2.3 A Trinucleotide repeat sequence

Fifty-seven base pairs downstream of the translation stop codon, starting at position 2438, is a series of nine CAG repeats. This sequence, together with the ensuing 21 base pairs, would, upon translation, yield an "OPA" motif (Wharton *et al.*, 1985) consisting of 14 glutamine residues interrupted by two histidines in positions 11 and 12 in the 16 residue sequence. Imperfect trinucleotide repeats such as these are common in *D. melanogaster* (Magoulas and Hickey, 1992) and other species (Tautz and Renz, 1984).

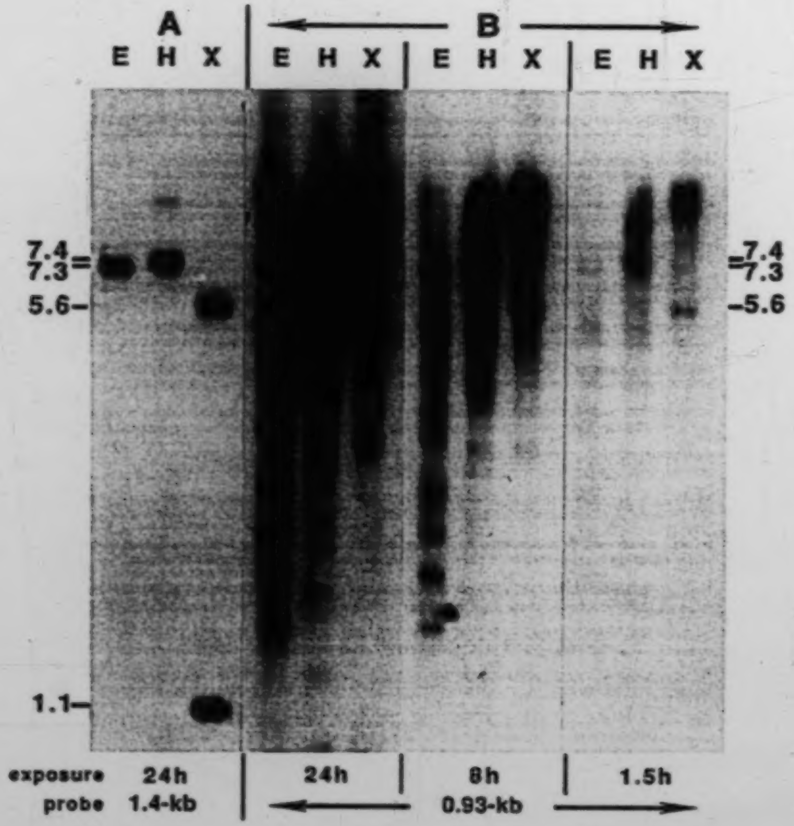
There are differences in the restriction pattern of the cDNAs, so that the second longest one, cDNA83, was sequenced, although only from one strand. In comparison with cDNA29, cDNA83 lacks 161 base-pairs at its 5′-end. Interestingly, it has seven CAG repeats compared to the nine found in cDNA29. The situation is reminiscent of variation in similar sequences in humans, where several heritable disease states have been shown to be correlated with instability leading to spontaneous amplification of trinucleotide repeats (Mandel, 1993). There is no evidence that similar mutants are to be found in the fruit-fly, nor even, of course, that the cDNA variant derived directly from genomic variation.

Nash *et al.* (1994) demonstrated that the repeat sequence causes significant background hybridization due, presumably, to similar trinucleotide repeats in other parts of the genome (Figure 3-10). Indeed, a more extensive, though rather disorganized, set of the same trinucleotide repeat is found in the ~2.0kb *Sau* 3A-*Eco* RI DNA segment upstream of the *ras* locus. Whether there is any functional relationship between the two repeats is unknown.

3.2.4 Inosine monophosphate dehydrogenase (IMPDH)

Conceptual translation of the long open reading-frame in the cDNA generates a polypeptide of 537 amino acids that is similar to the sequence found in all known IMPDH's (Table 3-3). The alignments between the *D. melanogaster* sequence and the IMPDH sequences found in current protein sequence libraries that are most (Human IMPDH 1) and least (*E. coil*) similar to it are shown in Figure 3-11. The

Figure 3-10 "OPA" sequence hybridization pattern



Hybridization of ^{32}P-labeled fragments from cDNA29 to *Eco* RI (E), *Hin* dIII (H), and *Xho* I (X) digests of genomic DNA from wild-type adult females. All four predicted signals are present in A; a weak high molecular weight signal seen in the *Hin* dIII digest is of unknown origin. In B, the three predicted signals from digestion of the *ras* DNA are identifiable in one or other autoradiograph since the 0.93kb probe is not expected to hybridize to the 1.1kb *Xho* I fragment.

A. Probed with a 1.4kb fragment from the 5′ end of cDNA29; long autoradiographic exposure.

B. An identically prepared autoradiograph, except the probe used was a 0.93kb fragment from the 3′ end of the cDNA, which contains an imperfect trinucleotide repeat (OPA) sequence; three different autoradiographic exposures. Background hybridization (seen in Figure 3-10) is effectively eliminated in A but greatly amplified in B.

(From Nash *et al.*, 1994)

Table 3-3 Similarity of the known IMPDH amino acid sequences

Species		% matched	Access number
Homo sapiens (1)	(human)	66	P20839
Mus musculus (1)	(mouse)	66	P50096
Homo sapiens (2)	(human)	65	P12268
Mesocricetus auratus (2)	(golden hamster)	65	P12269
Mus musculus (2)	(mouse)	65	P24547, Q61734
Pneumocystis carinii		57	Q12658
Saccharomyces cerevisiae	(baker's yeast)	56	P38697
Trypanosoma brucei		50	P50098
Leishmania donovani		49	P21620
Arabidopsis thaliana	(mouse-ear cress)	45	P47996
Acinetobacter calcoaceticus		38	P31002
Borrelia burgdorferi	(lyme disease spirochete)	38	P49058
Bacillus subtilis		37	P21879
Escherichia coli		37	P06981, P76574, P78202
Helicobacter pylori	(campylobacter pylori)	37	P56088
Haemophilus influenzae		35	P44334
Methanococcus jannaschii		35	Q59011
Mycobacterium leprae		34	Q49729
Mycobacterium tuberculosis		34	Q50715
Streptococcus pyogenes		34	P50099
Pyrococcus furiosus		32	P42851
Tritrichomonas foetus		29	P50097

The table shows the levels of amino acid sequence identity between IMP dehydrogenase from *Drosophila melanogaster* and other species. The amino acid sequences were found in the Swiss-Protein library. The access number for each sequence is included. The percentage of matched residues was obtained from pairwise alignments generated using the Clustal V program (Higgins and Sharp 1988) and includes only identical matches. Sample alignments produced by this program are shown in Figure 3-11. (Adapted from Nash *et al.*, 1994)

Figure 3-11 Comparison of IMPDH sequences

```
E  MQSVTLCIMPRQYLLTTLVEILPHLRIAK---------------------EALTFDDVLLVPAHSTVLPN  49
   |:| |  :    :: :|     | ::                        |:||::| |::|:     ::
D  MESTTK-VKVNGFVESTSSSAAPAIQTKSTTGFDAELQDGLSCKELFQNGEGLTYNDFLILPGYIDFTAE  69
   |     /    ::: | :            ||: :|  |||: ::|| :::||||||||||||:||| |:
H  MA----/----DYLISGG------------TGYVPE--DGLTAQQLFASADDLTYNDFLILPGFIDFIAD  48

E  TADLSTQLTKTIRLNIPMLSAAMDTVTEARLAIALAQEGGIGFIHKNMSIERQAEEVRRVKKHESGVVTD 119
     |||: |||:: |  |::|::||||||: :|||:|  ||||:|| | : | || ||::|||   | : |
D  EVDLSSPLTKSLTLRAPLVSSPMDTVTESEMAIAMALCGGIGIIHHNCTPEYQALEVHKVKKYKHGFMRD 139
   ||||:|:||: :||::||:||||||||||::|||||| ||||:||||||||:|| ||:|||:: :||: |
H  EVDLTSALTRKITLKTPLISSPMDTVTEADMAIAMALMGGIGFIHHNCTPEFQANEVRKVKNFEQGFITD 118

E  PQTVLPTTTLREVKELTERNGFAGYPVVTEEN---ELVGIITGRDVRF/VTDLNQPVSVYMTPK//ERLV 183
   |  : || |: :| |   :|||:||||  : :    |:|::|:||: |/  :  : :   :   //| ||
D  PSVMSPTNTVGDVLEARRKNGFTGYPVTENGKLGGKLLGMVTSRDIDF-RENQPEVLLADIMTT--E-LV 205
   | |:||::||||||||: ::||:| |:|| | :|:||:|:||||||||  |::   ||:::||   |/||
H  PVVLSPSHTVGDVLEAKMRRHGFSGIPITETGTMGSKLVGIVTSRDIDFLAEKDHTTLLSEVMTPRIE/LV 187

E  TVREGEAREVVLAKMHEKRVEKALVVDDEFHLIGMITVKDFQKAEAKPNACKDEQGRLRVGAAVGAGAGN 253
   |  :|       | ::  :  |  :|::  :|::||:  |::|| : ||| || : :| ||||:|: : :
D  TAPNGINLPTANAILEKSKKGKLPIVNQAGELVAMIARTDLKKARSYPNASKDSNKQLLVGAAIGTRSED 275
    || |: |  || ||::||||||||||: :|||:||||||||| | || |||||:|||| |||:||| :|
H  VAPAGVTLKEANEILQRSKKGKLPIVNDCDELVAIIARTDLKKNRDYPLASKDSQKQLLCGAAVGTREDD 257

E  EERVDALVAAGVDVLLIDSSHGHSEGVLQRIRETRAKYPDLQIIGGNVATAAGARALAEAGCSAVKVGIG 323
     |:  ||| ||||:::|||:|:|   :: |:  :  ||:||:||||||  | |: | :||  :::||:|
D  KARLALLVANGVDVIILDSSQGNSVYQVEHIKYIKETYPELQVIGGNVVTRAQAKNLIDAGVDGLRVGMG 345
   | || ||   |||||:|||||||||||: |: |||: ||:|||||||||| |||||||||||||||||||
H  KYRLDLLTQAGVDVIVLDSSQGNSVYQIAMVHYIKQKYPHLQVIGGNVVTAAQAKNLIDAGVDGLRVGMG 327

E  PGSICTTRIVTGVGVPQITAVADAVEALEGTGIPVIADGGIRFSGDIAKAIAAGASAVMVGSMLAGTEES 393
   :|||| |: | : | || ||| :        |:||||||||:  |:| |||| ||||||:||:|||| |:
D  SGSICITQEVMACGCPQATAVYQVSTYARQFGVPVIADGGIQSIGHIVKAIALGASAVMMGSLLAGTSEA 415
    ||||||||||||| ||:||||:|: |||:||||:|||||||::||:|||:|||||:||||||||:|:||
H  CGSICITQEVMACGRPQGTAVYKVAEYARRFGVPIIADGGIQTVGHVVKALALGASTVMMGSLLAATTEA 397

E  PGEIELYQGRSYKSYRGMGSLGAMSKGSSD-----RYFQSDNAADKL-VPEGIEGRVAYKGR-LKEIIHQ 456
   |||  : :|   | ||||||| || :| :      ||::::   ||: |::|: | :  ||  |: :
D  PGEYFFSDGVRLKKYRGMGSLEAMERGDAKGAAMSRYYHNE--MDKMKVAQGVSGSIVDKGSVLRYLPYL 483
   ||||||||||||||||||||||:|||::  :::  ||: :|// ||:|:|||||||| ||||: :::|||
H  PGEYFFSDGVRLKKYRGMGSLDAMEKS---SSSQKRYF-SE//GDKVKIAQGVSGSIQDKGSIQKFVPYL 461

E  QMGGLRSCMGLTGCGTIDELRTKA-----EFVRISGAGIQESHVHDVTITKESPNYRLGS  516
   : |  :|| :: |  :|: ||         |:: : ::  |::||:: :: |   :
D  ECGLQHSCQDI-GANSINKLRDMIYNGQLRFMKRTHSAQLEGNVHGL-FSYEKRLF      537
     |:||:||||/|| |:: || |:|:|:|:| ||| |||:|| ||||/ ||||||:
H  IAGIQHGCQDI/GARSLSVLRSMMYSGELKFEKRTMSAQIEGGVHGL/HSYEKRLY      515
```

Comparison of the conceptual translation product from the *Drosophila melanogaster ras* locus (D, middle sequence) with IMP dehydrogenase from *Escherichia coli* (E, above) and *Homo sapiens* (H, below). Two independent comparisons to the *Drosophila melanogaster* sequence are shown, so the alignment of the *E. coli* and *H. sapiens* sequences is not necessarily optimal. Colons indicate conservative substitutions (score 8 or more in PAM250). Hyphens indicate gaps generated by CLUSTAL V to optimize alignment. Solidi indicate gaps that bring the three sequences into register for display purposes but have no statistical or evolutionary significance.

(From Nash *et al.*, 1994)

The probable identity of this conceptual translation product of the *ras* locus was previously reported by Nash and Hu (1992) and Nash *et al.* (1994). Slee and Bownes (1993) have reported, also in abstract, the existence of a similar sequence, similarly identified as IMPDH, associated with an anonymous gene. It was identified in an enhancer-trap screen for testis-specific enhancers and its transcription is regulated temporally and in a tissue specific manner. They observed several distinct RNA transcripts by using the IMPDH cDNA probe.

3.3 The molecular lesions in *ras* mutant DNA

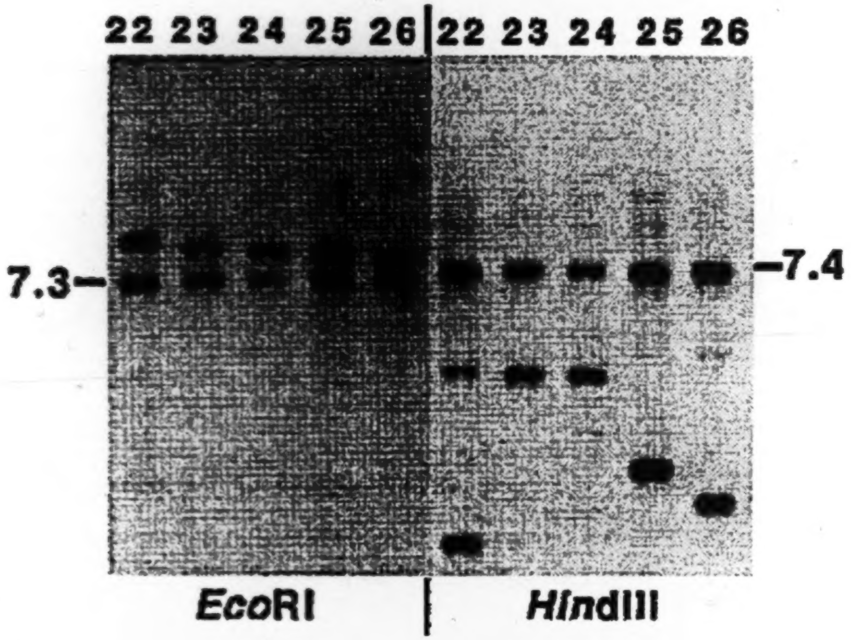
3.3.1 Southern analysis

The initial stages of the analysis of the molecular lesions is already published (Nash *et al.*, 1994), but is reported here as an integral part of the somewhat more complete analysis which is now available.

All five dysgenically induced *ras* lethal mutants (*ras*$^{l22\text{-}l26}$, or NJ1, 4, 5, 6 and 8, Figure 3-12) and the four spontaneous viable *ras* mutants (*ras*$^{1\text{-}4}$, Figure 3-13) showed DNA alterations in the 4.0kb *Eco* RI-*Hin* dIII region (Nash *et al.*, 1994). In Figure 3-12, genomic DNA of heterozygous females with the P-element induced *ras-lethal* chromosome and the *FM6* balancer chromosome was digested by *Eco* RI and *Hin* dIII. Southern analysis shows that a common wild-type 7.3kb *Eco* RI fragment and a 7.4kb *Hin* dIII fragment are present in all five strains. These two wild-type fragments are derived from the *FM6* chromosome. The *Eco* RI digestion pattern shows that each mutant strain contains an additional large signal, indicating that there probably is no *Eco* RI site in the P-element insert and the size of the insert is about from 1.0 to 2.0kb. In the *Hin* dIII digestion, there are additional smaller bands present in all five strains, suggesting that there is at least one *Hin* dIII site in each insert. Considering the size of these additional *Hin* dIII fragments and their hybridization signal, the insertion site in each of these P-element induced mutants probably is closer to the *Eco* RI site than to the *Hin* dIII site of the 4.0kb *Eco* RI-*Hin* dIII fragment.

Southern analysis indicated that, in three of the four spontaneous visible mutants, *ras*^{1}, *ras*^{2} and *ras*^{4}, the most likely molecular lesion is an insertion within the 1.1kb *Xho* I-*Xho* I region found in the 4.0kb *Eco* RI-*Hin* dIII fragment of the wild-type DNA (Figure 3-13). In *ras*^{2}, there is a single 6.0kb *Xho* I-*Xho* I fragment which replaces the original fragment, suggesting an approximately 5kb insert, with no *Xho* I sites. In *ras*^{1} and *ras*^{4} the wild-type 1.1kb *Xho* I-*Xho* I fragment is replaced by two fragments (4.2kb and 1.1kb in *ras*^{1}, 4.7kb and 3.6kb in *ras*^{4}) which hybridize to the 1.1kb probe. However, a single band was found in the Southern analysis with *ras*^{1} digested by *Bam* HI and *Sal* I and *ras*^{4} digested by *Bam* HI, indicating both of them are insertion mutants rather than rearrangements. The digestion pattern suggests that two different inserts each containing at least one *Xho* I site are inserted into *ras*^{1} and *ras*^{4}, respectively. To study the nature of the inserts, especially the insertion sites, the five insert containing segments were cloned.

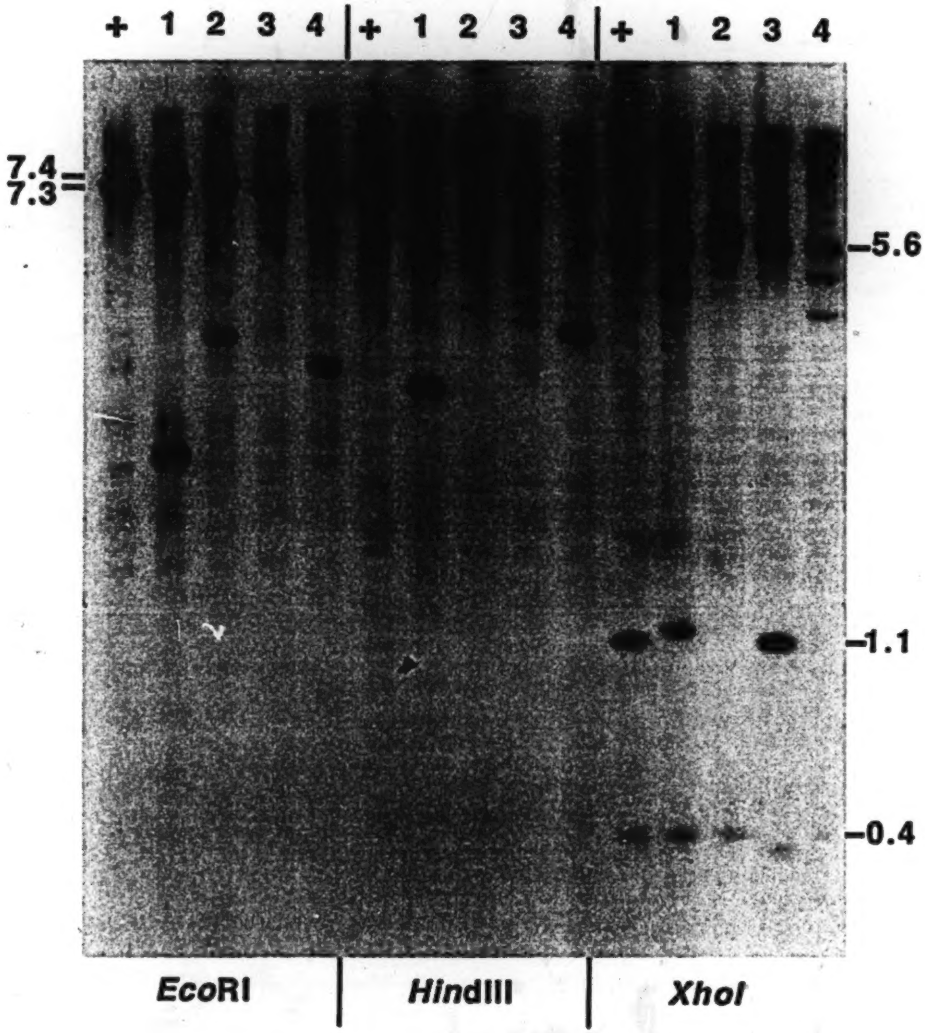
Figure 3-12 Southern analysis of P-element induced lethal mutants



Hybridization of ^{32}P-labeled 4.0kb *Eco* RI-*Hin* dIII DNA to digests of genomic DNA from adult females carrying five different *raspberry-lethal* mutants generated in P-M mutagenesis screens (*ras*$^{l22-l26}$, 22-26). All flies were heterozygous at the *ras* locus, being *FM6(ras*$^{+}$*)/rasl*. Constant-sized bands represent the wild-type restriction fragments from FM6; variable restriction fragments are derived from the mutants.

(From Nash *et al.*, 1994)

Figure 3-13 Southern analysis of Am$^+$ and *ras* viable mutants



Hybridization of ^{32}P-labeled 4.0kb *Eco* RI-*Hin* dIII DNA to Southern-transferred digests of genomic DNA from wild-type (+) and *ras1-4* (1-4) adult females. All flies are homozygous at the *ras* locus. The size marks indicate, on the left, the wild-type *Eco* RI (7.3kb) and *Hin* dIII (7.4kb) fragments and, on the right, the visible wild-type *Xho* I fragments. Long exposure used to visualize the 0.4kb *Xho* I fragment reveals low-level hybridization elsewhere on the autoradiograph. The origin of this background signal is clarified in Figure 3-10. The background obscures potential visualization of an ~11kb *Xho* I fragment (with ~0.2kb of homology with the 4.0kb probe) located upstream of the locus. The fragment is unaltered in the mutants.

(From Nash et al., 1994)

3.3.2 Cloning of the insertion point containing fragments in *ras*^{1}, *ras*^{2} and *ras*^{4}

3.3.2.1 *ras*^{1}

The initial attempt was to clone the *Xho* I-*Xho* I inserts directly into *Xho* I digested pBluescript plasmid. The inserts were gel purified in large quantity. Dot blot hybridization confirmed that all five purified DNA samples hybridized to the 1.1kb *Xho* I-*Xho* I fragment. The prepared inserts were ligated to *Xho* I digested pBluescript DNA. However, almost all the clones recovered were religated plasmids even through the *Xho* I digested pBluescript plasmid was BAP (bacterial alkaline phosphatase) dephosphorylated. After several rounds of ligation and plating, several clones were apparently identified by hybridization to the 1.1kb *Xho* I-*Xho* I fragment, but all of them were false positives.

To facilitate recovery of the desired ligation products and reduce the background ligation, a new approach was taken. The enzyme *Sal* I produces the same four base cohesive ends as *Xho* I does (TCGA). If there is no *Sal* I site present in the insert, the purified *Xho* I digested DNA fragments, together with the insert, could be ligated into the plasmid's *Sal* I cloning site. This will destroy both the *Xho* I and *Sal* I sites by the *Xho* I-*Sal* I religated sequences. The ligated products could be then digested by *Sal* I before transformation. This would dramatically reduce the frequency of the religated plasmids and the frequency of recombinant plasmids bearing *Sal* I sites in the insert DNA. Southern analysis revealed that there is no *Sal* I sites in *ras*^{1}. Thus both *Xho* I insert fragments of *ras*^{1} were cloned by using this approach.

3.3.2.2 *ras*^{2}

The single *Xho* I insert fragment of *ras*^{2} was also cloned by using the above approach since there is no *Sal* I site in it.

3.3.2.3 *ras*^{4}

The *Xho* I-*Sal* I ligation method cannot be used to clone *ras*^{4} insert since there is at least one *Sal* I site in the insert. The final *Sal* I digestion of the ligated products will destroy the recombinants at the insert's *Sal* I site. Searching for other combinations of double digestion revealed that *Xho* I-*Eco* RI double digestion generated two fragments (2.7kb and 1.6kb) and both of them hybridized to the 1.1kb probe. These two *ras*^{4} fragments were cloned into pBluescript's *Xho* I-*Eco* RI cloning site. This approach would not only reduce the plasmid self ligation but also fix the orientation of the insert within the plasmid.

Thus, the wild type 1.1kb *Xho* I-*Xho* I fragment along with its adjacent inserts in each mutant was cloned into plasmids. In *ras*^{2}, the whole insert was cloned into a single plasmid. In *ras*^{1}, the DNA fragments from both *Xho* I sites inside the insert to the *Xho* I sites in the adjacent wild-type region were cloned into two different plasmids. In *ras*^{4}, the DNA fragments from both *Eco* RI sites inside the insert to the *Xho* I sites in the adjacent wild-type region were cloned into two different plasmids.

3.3.3 Analysis of the insertion points

3.3.3.1 *ras*^{1}

All five plasmids were initially sequenced from their genomic *Xho* I sites into the cloned DNA and towards the insertion sites. Preliminary analysis of the sequencing gels indicated that all three insertions occurred within roughly 0.4kb from the same end of the 1.1kb segment and clearly within the region that seems to be transcribed. In all cases, the insertion target regions were sufficiently remote from the plasmid primer site that accurate sequence reading was not possible. To pinpoint the insertion sites, two sequencing primers, Primer3 and Primer4, located on either side of, and directed towards, the presumptive insertion were synthesized (*Section 2.4.10*).

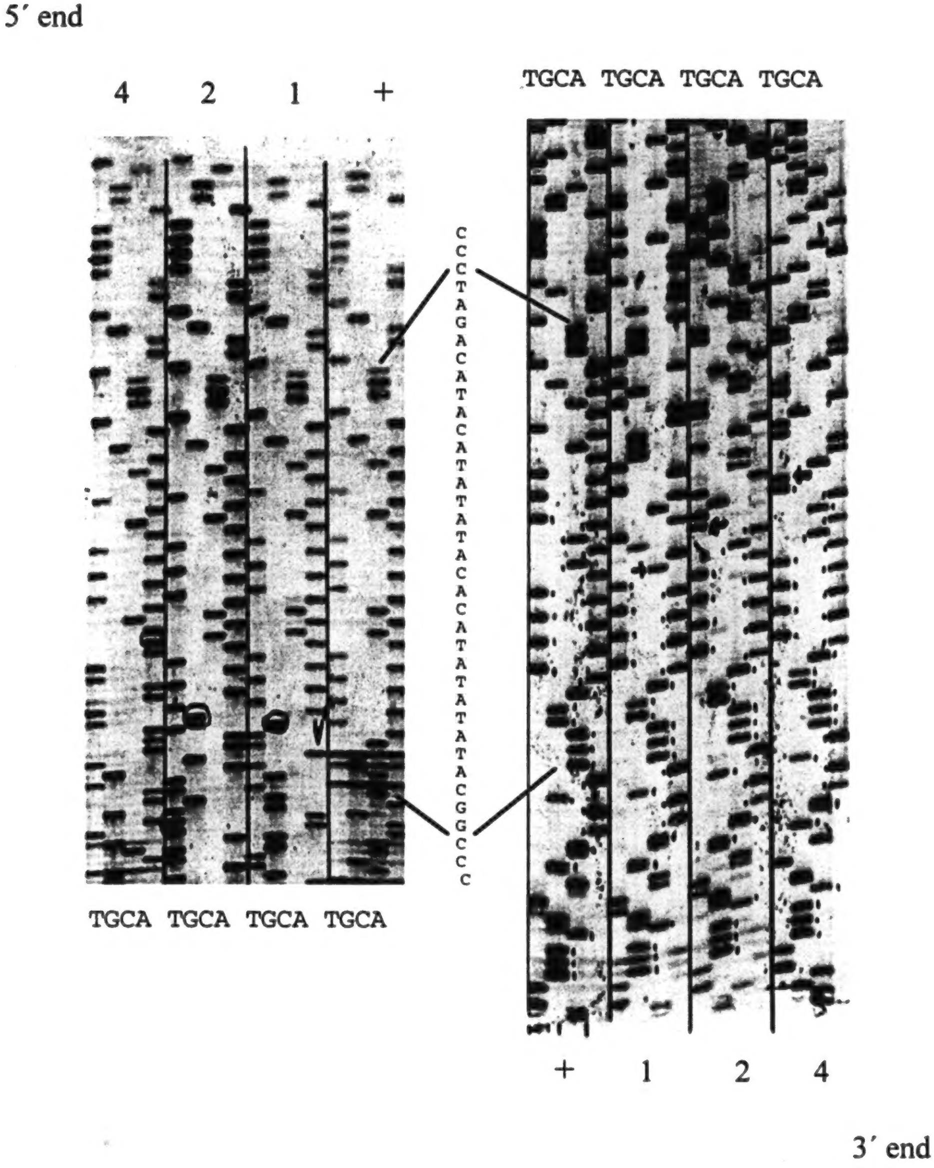
Sequencing from these two primers revealed that the *ras*^{1} insert was inserted in a poly-AT rich sequence in the second intron. There is a 6bp duplicated sequence in the insertion site (Figure 3-14, Figure 3-15). The two insert fragments were sequenced to the maximal resolution from both internal and external ends. The four sequences were sent to Genbank to conduct a BLAST search. No similar DNA sequence or functional domain was found. Sequencing comparison between the two external ends of the insert showed that the insert contains a direct long terminal repeat (LTR) at both ends (Figure 3-16).

3.3.3.2 *ras*^{2}

The *ras*^{2} insert was inserted in the same poly-AT rich sequence as the *ras*^{1} insert with an 8bp duplicated sequence (Figure 3-14, Figure 3-15). The ends of the *ras*^{2} insert match the ends of "blastopia polyprotein" (BLPP), a micropia-like transposon element (Frommer *et al.*, 1994). Sequence comparison revealed that the 5′ end of the *ras*^{2} insert was 96.3% identical to the 3′ end of BLPP, and the 3′ end of the *ras*^{2} insert was 94.5% identical to the 5′ end of BLPP except the first 370bp of BLPP's 5′ end was missing in the 3′ end of the *ras*^{2} insert (Figure 3-17). The LTRs in both cases, which are part of the ends, are almost identical.

There are four *Eco* RI sites in BLPP, yielding three internal fragments of 1.1kb, 0.2kb and 70bp and two external fragments (1.8kb and 2.2kb), which also contain genomic DNA. Only four *Eco* RI fragments were found in the *ras*^{2} insert bearing pBluescript, 0.2kb, 1.1kb, 2.0kb and 6kb. The 0.2kb and 1.1kb *ras*^{2} fragments should be equivalent to the 0.2kb and 1.1kb BLPP fragments located in the middle of the insert. No 70bp BLPP fragment was seen in the Ethidium bromide stained gel, but this fragment is present (See below). To further analyze the insert, the *ras*^{2} insert bearing pBluescript was partially digested by *Eco* RI and religated. Thirty plasmids were selected. *Eco* RI digestion showed that there were mainly of three kinds, 6kb, 7.1Kb and 8.0Kb. Two of the 6kb, four of the 6 + 1.1kb and four of the 6 + 2kb were randomly chosen and sequenced from the ends adjacent to the universal primer. The results revealed six different patterns, including one containing two adjacent *Eco* RI fragments and one containing two noncontiguous DNA fragments of the insert. These sequences revealed that the *ras*^{2} insert is a micropia-like, BLPP transposon (Figure 3-18; Table 3-4).

Figure 3-14 Sequence analysis of *ras* insert points



Sequencing from both directions on the insertion sites in *ras*^{1}(1), *ras*^{2} (2) and *ras*^{4} (4) along with the wild-type 4.0kb *Eco* RI-*Hin* dIII fragment (+). The insertion point is either marked by a circle (5′ end) or an arrow (3′ end). In the middle of the two figures is the wild-type sequence spanning the insertion sites.

Figure 3-15 Diagram of the insertion sites of *ras*^{1}, *ras*^{2} and *ras*^{4}

A

```
5´
    GCTGGTTAGT TCGCCCATGG ACACGGTAAC CGAATCGGAG ATGGCCATCG  399
                                   primer shu 4 »»
    CCATGGCGgt aagtatcgga ttttaatgaa tccctagaca tacatatata  449

    cacatatata tacggcccag aaccgctggc gggttcgcct gagctgtgct  499
   insertion site
    agcctgtggt tgtaactcaa cactcttatg cgcctctcta accgccaatg  549
                     «« primer shu3
    tttctatcca ctgctcacac accatccaaa tccattcgaa tcccgcgcgc  599

    agCTGTGTGG TGGCATTGGC ATCATCCATC ACAACTGCAC GCCGGAATAC  649
                                                                      3´
```

This diagram is to show the second intron of IMPDH gene (labeled), the two primers used in the sequencing of *ras* inserts (underline) and the 12bp region which is the target for all three inserts (underline).

B

ras^1 insert, minimum size 4.2kb, terminal duplication 6bp.

```
(+)         5'  ccctagacatacatatatacacatatatatacggccc
(1)                    catacatatatacacatatata GTTA
                                  TGAC tatatatacggcccag   3'
```

ras^2 insert, minimum size 5.0kb, terminal duplication 8bp.

```
(+)         5'  ccctagacatacatatatacacatatatatacggccc
(2)                    catacatatatacacatatat GTAA
                               TGTT acatatatatacggcccag   3'
```

ras^4 insert, minimum size 4kb, terminal duplication 4bp.

```
(+)         5'  ccctagacatacatatatacacatatatatacggccc
(4)                    catacatatataca AATT
                            AACT tacacatatatatacggcccag   3'
```

Insertion sites of *ras*^{1} (1), *ras*^{2} (2) and *ras*^{4}(4). This diagram is generated from the data in Figure 3-14. (+) = wild-type sequence. Capital letter = internal insert sequence.

Figure 3-16 Comparisons of the termini of the insertion sequence in the *ras*$^{1, 2 \text{ and } 4}$ mutants

All three *ras* inserts contain direct long terminal repeats. The lfasta alignments show the similarity of the terminal repeats of each individual insert. The 3′ end sequences used in the alignments are the anti-parallel sequence of the actually sequenced direction. There is no significant similarity between the terminal repeats of any two of these inserts.

5 = 5′ end
3 = 3′ end

ras^{1}

95.4% identity in 131 nt overlap

```
              140        150       160       170       180
ras1-5 CTTTGTCGATTAAAGA-GGGGTTAAAACAAAACGAAAGGTTTCGCTTAACAAATTGAGGG
       :::::::::::::::: :::::::::::::::::::::::::::::::::::::::::::
ras1-3 CTTTGTCGATTAAAGAGGGGGTTAAAACAAAACGAAAGGTTTCGCTTAACAAATTGAGGG
               10        20        30        40        50        60

     190       200       210       220       230         240
ras1-5 TTTATTCAATGATGTGTACAGATCGAGGTCTTCTTGTCTACTGACTAAAA--TAAGCTAA
       ::::::::::::::::::::::::::::::::::::::::::::::::::   :::::
ras1-3 TTTATTCAATGATGTGTACAGATCGAGGTCTTCTTGTCTACTGACTAAAAATAAAGCTTT
               70        80        90       100       110       120

       250
ras1-5 GCTCAGACAAA
       :::::::::::
ras1-3 GCTCAGACAAA
              130
```

ras^{2}

99.3% identity in 274 nt overlap

```
               10        20        30        40        50
ras2-5 GTAAGATTGTTTATTATTATTGTTTATTATTAATTTAATTATTAGTTTAATTCTAAATG-
       :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
ras2-3 GTAAGATTGTTTATTATTATTGTTTATTATTAATTTAATTATTAGTTTAATTCTAAATGA
          130       140       150       160       170       180

      60        70        80        90        100       110
ras2-5 CGGTAATTATATAAAGTATTCTTAGTTTGTGACC-GAACGACTGCTTCTGAGAACTCGAC
       :::::::::::::::::::::::::::::::::: :::::::::::::::::::::::::
ras2-3 CGGTAATTATATAAAGTATTCTTAGTTTGTGACCGGAACGACTGCTTCTGAGAACTCGAC
          190       200       210       220       230       240

120       130       140       150       160       170
ras2-5 TGACTGCAACGCTGACTTCCACTCGACTCTACTGCTGCCTTGCGATCGTTCCTTCGATGA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
ras2-3 TGACTGCAACGCTGACTTCCACTCGACTCTACTGCTGCCTTGCGATCGTTCCTTCGATGA
          250       260       270       280       290       300

180       190       200       210       220       230
ras2-5 CTGATTGTCGATAACTTCGATTTCCGACAATCATGTTGCTGCTCAACAGCGCTGCCAACC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
ras2-3 CTGATTGTCGATAACTTCGATTTCCGACAATCATGTTGCTGCTCAACAGCGCTGCCAACC
          310       320       330       340       350       360

      240       250       260       270
ras2-5 GGGTTGTTGCCAGCCTTCAGCCTTACTCCATGTT
       ::::::::::::::::::::::::::::::::::
ras2-3 GGGTTGTTGCCAGCCTTCAGCCTTACTCCATGTT
          370       380       390
```

ras^{4}

93.2% identity in 307 nt overlap

```
              110       120       130        140       150
ras4-5 AGACCTCTTCTCCTAGATTGGGAGATATGGT-GGGAGAACGTCTCTCCGTTG-TTGACTG
       ::::::::::::::::::::::::::::::: :::::::::::::::::::: :::::::
ras4-3 AGACCTCTTCTCCTAGATTGGGAGATATGGTGGGGAGAACGTCTCTCCGTTGTTTGACTG
           10        20        30        40        50        60

      160       170       180       190       200       210
ras4-5 CCCTTAAGGCTAGCCAACCAATTCAATGATAACAGCAGTAGCTGGAGTTGATTGAAGGCG
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::
ras4-3 CCCTTAAGGCTAGCCAACCAATTCAATGATAACAGCAGTAGCTGGAGTTGATTGAA---G
           70        80        90       100       110       120

      220        230       240       250       260       270
ras4-5 CGATGCG-TCTTTATTGAATACAAAATCAAACTGACTATAAGCTAACAAAGGGAAAAACA
       ::::::: :::::::::::::::::::::::  ::::::::::::::::: :::::::::
ras4-3 CGATGCGCTCTTTATTGAATACAAAATCAAA--GACTATAAGCTAACAAA-GGAAAAACA
             130       140       150         160        170

       280       290       300         310       320       330
ras4-5 TCATAGCGGCCCTCTGCCAATGCGCAGA--GCTTCTGCCGGCTATGCATGAGCTTCACGG
       ::::::::: ::::::::::: ::::::    :::::::::::::::::::::::: :::
ras4-3 TCATAGCGG-CCTCTGCCAAT-CGCAGACGATTTCTGCCGGCTATGCATGAGCTTC-CGG
      180        190        200       210       220       230

         340       350       360       370       380         390
ras4-5 CCAAATGCTTGGTCAGCAATTTGACCGGTGCTGGTGTTGACGACGATCAGTC--GTTAAC
       :::::::::::::::::::::::::::::::::::: ::  :::::::::::  ::::::
ras4-3 CCAAATGCTTGGTCAGCAATTTGACCGGTGCTGGTG-TGCGGACGATCAGTCCGGTTAAC
         240       250       260       270        280       290

           400
ras4-5 TTAGTTA
       :::::::
ras4-3 TTAGTTA
          300
```

Figure 3-17 Sequences of the 5′ LTR of the *ras*^{2} insert and the 3′ termini of the transposon blastopia polyprotein (BLPP)

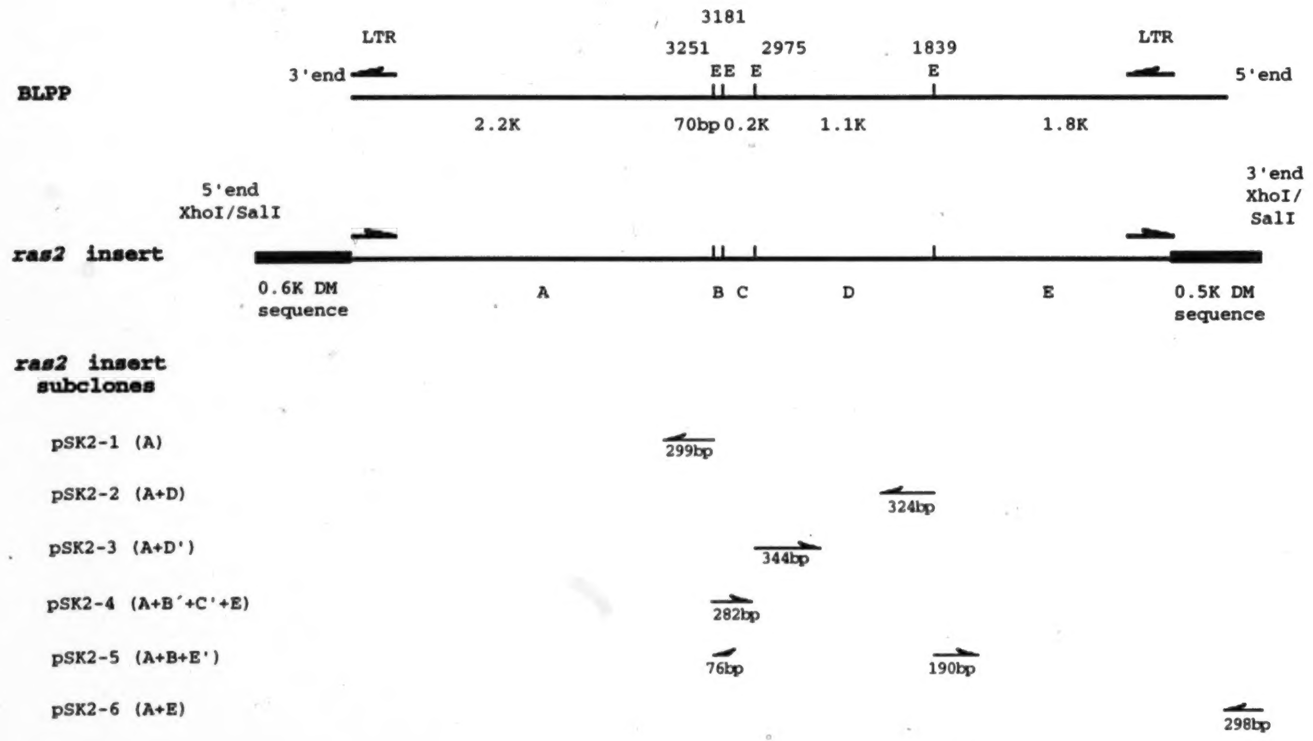
A *ras*^{2} 5′ end terminal repeat (anti-parallel) and BLPP 3′ end terminal repeat (97.1% identity in 279 nt overlap)

```
     370        380       390       400       410       420
DMBLPP AATGTAACAT-GAGTAAGGCTGAAGGCTGGCAACAACCCGGTTGGCAGCGCTGTTGAGCA
       :::::::::: :::::::::::::::::::::::::::::::::::::::::::::::::
ras2-5 AATGTAACATGGAGTAAGGCTGAAGGCTGGCAACAACCCGGTTGGCAGCGCTGTTGAGCA
        130       140       150       160       170       180

      430       440       450       460       470       480
DMBLPP GCAACATGATTGTCGGAAATCCAAGTTATCGACAATCAGTCATCGAAGG-ACGATCGC-A
       ::::::::::::::::::::: ::::::::::::::::::::::::::: :::::::: :
ras2-5 GCAACATGATTGTCGGAAATCGAAGTTATCGACAATCAGTCATCGAAGGAACGATCGCAA
        190       200       210       220       230       240

        490       500       510       520       530       540
DMBLPP GGCAGCAGTAGAGGCGAGTGGAAGTCAGCGTTGCAGTCAGTCGAGTTCTCAGCAGCAGTC
       ::::::::::::: :::::::::::::::::::::::::::::::::::::: :::::::
ras2-5 GGCAGCAGTAGAGTCGAGTGGAAGTCAGCGTTGCAGTCAGTCGAGTTCTCAGAAGCAGTC
        250       260       270       280       290       300

        550       560       570       580       590       600
DMBLPP GTTCGGTCCACAAACTAAGAAATACTTTATATAATTACCGCATTTAGAATTAAACTAATA
       ::::::: ::::::::::: ::::::::::::::::::::::::::::::::::::::::
ras2-5 GTTCGGT-CACAAACTAAG-AATACTTTATATAATTACCGCATTTAGAATTAAACTAATA
        310        320        330       340       350       360

        610       620       630       640
DMBLPP ATTAAATTAATAATAAACAATAATAATAAACAATCTTAC
       :::::::::::::::::::::::::::::::::::::::
ras2-5 ATTAAATTAATAATAAACAATAATAATAAACAATCTTAC
          370       380       390       400
```

B *ras*^{2} 3′ end terminal repeat and BLPP 5′ end terminal repeat (96.4% identity in 277 nt overlap)

```
             1140      1150      1160      1170      1180      1190
DMBLPP AACATGGAGTAAGGCTGAAGGCTGGCAACAACCCGGTTGGCAGCGCTGTTGAGCAGCAAC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
ras2-3 AACATGGAGTAAGGCTGAAGGCTGGCAACAACCCGGTTGGCAGCGCTGTTGAGCAGCAAC
               10        20        30        40        50        60

             1200      1210      1220      1230      1240      1250
DMBLPP ATGATTGTCGGAAATCGAAGTTATCGACAATCAGTCATCGAAGGAACGATCGCAAGGCAG
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
ras2-3 ATGATTGTCGGAAATCGAAGTTATCGACAATCAGTCATCGAAGGAACGATCGCAAGGCAG
               70        80        90       100       110       120

             1260      1270      1280      1290      1300      1310
DMBLPP CAGTGGAGTAGGAGTGGAAGTCAGCGTTGCAGTCAGTCGTGTTCTCAGCAGCAGTTCGTT
       :::: ::::  :::::::::::::::::::::::::::: :::::::: ::::::   :
ras2-3 CAGTAGAGT-CGAGTGGAAGTCAGCGTTGCAGTCAGTCGAGTTCTCAGAAGCAGTCGTTC
               130       140       150       160       170

             1320      1330      1340       1350      1360
DMBLPP CGGTCACAAACTAAGAATACTTTATATAATTACCG-CATTTAGAATTAAACTAATAATTA
       ::::::::::::::::::::::::::::::::::: ::::::::::::::::::::::::
ras2-3 CGGTCACAAACTAAGAATACTTTATATAATTACCGTCATTTAGAATTAAACTAATAATTA
     180       190       200       210       220       230

    1370      1380      1390      1400
DMBLPP AATTAATAATAAACAATAATAATAAACAATCTTACAT
       :::::::::::::::::::::::::::::::::::::
ras2-3 AATTAATAATAAACAATAATAATAAACAATCTTACAT
     240       250       260       270
```

Figure 3-18 Diagram of the sequence similarity between the *ras*^{2} insert and the BLPP



Similarity comparison between *ras*^{2} insert and BLPP. The position of *Eco* RI sites in BLPP is adapted from the BLPP sequence. The restriction map of the *ras*^{2} insert is generated from above sequence data. The dark arrows above the map represent the LTRs on both BLPP and *ras*^{2} insert. The diagram shows the partial sequences from *Eco* RI sites of the subclones of the *ras*^{2} insert. The arrows indicate the approximate sequenced regions and the sequence directions.

Table 3-4 Sequence comparison of internal segments of the *ras*^{2} insert with the BLPP sequence

Origin of the sequence	Sequenced length (base-pairs)	Equivalent to BLPP sequence	Number of *Eco* RI sites in the region
pSK2-1 (6.0kb)	299	3251-3547bp	1
pSK2-2 (7.1kb)	324	1839-2162bp	1
pSK2-3 (7.1kb)	344	2980-2637bp	1
pSK2-4 (8.0kb)	282	3256-2975bp	3
pSK2-5 (8.0kb) *	76	3181-3256bp	2
	193	1844-1648bp	1
pSK2-6 (8.0kb)	28 270 (total 298bp)	plasmid sequence DM IMP sequence	span the original cloning site Xho I/Sal I (CTCGAG/C)

The *ras*^{2} insert bearing plasmid contains four *Eco* RI sites which separate the plasmid into four fragments, 0.2kb, 1.1kb, 2.0kb and 6.0kb, on the agarose gel (a 70bp fragment was unable to be identified in the gel). The plasmid was partially digested by *Eco* RI followed by religation and transformation. Three kinds of plasmids, 6.0kb, 7.1kb and 8.0kb, were recovered. Based on the digestion pattern, the 6.0kb contains the plasmid plus about 3.0kb of the insert DNA, the 7.1kb contains the 6.0kb and 1.1kb *Eco* RI fragments and the 8.0kb contains the 6.0kb and 2.0kb fragments. The plasmids were sequenced from the universal primer and the data are shown here.

* plasmid pSK2-5 contains two noncontiguous *Eco* RI fragments.

3.3.3.3 *ras*^{4}

The *ras*^{4} insert was also inserted in the same poly-AT rich sequence as *ras*^{1} and *ras*^{2} with a 4bp duplicated sequence. As seen with both *ras*^{1} and *ras*^{2} inserts, the *ras*^{4} insert also contains a LTR end which is closely related to the LTR of a copia-like 8.8kb transposon (Figure 3-19, Karlik and Fyrberg, 1985). No DNA sequence and functional domain similar to the internal ends of the *ras*^{4} insert was found.

Sequencing data from all three *ras* inserts revealed that these are inserted in a 12bp region (Figure 3-14, Figure 3-15). All three inserts contain a short flanking duplication at the insertion site. This 12bp region and its surrounding sequence was sent to Genbank to conduct a BLAST search. No homologous sequence or functional domain was found.

All three inserts contain LTR sequences at both 5′ and 3′ ends. Sequence comparison showed that there was no relationship among LTR sequences of the different inserts (Figure 3-16).

3.3.4 Analysis of *ras*^{3} mutation

Unlike the other three spontaneous *ras* mutants, which show the mutant eye-colour in both males and females, the mutant eye-colour is present in *ras*^{3} males but not in *ras*^{3} females. Hemizygous females (*ras*^{3} over deficiency) show the identical mutant eye-colour as the hemizygous *ras*^{3} males, indicating, in addition to the eye-colour function, dosage compensation is abolished in *ras*^{3} (Nash *et al.*, 1994). Not surprisingly, Southern analysis revealed that a different kind of mutation, a deletion in the 447bp *Xho* I-*Xho* I region of the 4.0kb *Eco* RI-*Hin* dIII fragment, is present in *ras*^{3} (Figure 3-13).

To further investigate the mutation, the ~0.4kb *Xho* I-*Xho* I fragments of *ras*^{3} were gel purified and ligated to *Xho* I digested pBluescript. After several rounds of ligation and transformation, no positive clones were found, presumably due to the same problems as cloning the *ras* inserts. To clone this fragment, then, PCR amplification was used. Two primers (Primer1 and Primer2) were made just outside the *Xho* I sites. The *ras*^{3} fragment was amplified successfully and cloned into the pGEM-T plasmid. Sequencing both of the directly amplified products and the cloned plasmids revealed that there was a 37bp deletion (Figure 3-20). The deletion region is upstream of the potential transcriptional start site, and just in front of the P-element insertion site in *ras*^{l24}. The similar mutant location of *ras*^{l24} and *ras*^{3} suggests that this region presumably is related to eye-colour formation and dosage compensation, as well as a target for P-element mutagenesis.

The sequence CACCAC is present at the 5′ end and CAC at 3′ end of the deletion in the wild-type DNA; only a single CAC trinucleotide remains in the mutant DNA. The 37bp sequence and its adjacent sequences were sent to Genbank to conduct a BLAST search. No similar homologous sequence or functional domain was found.

Figure 3-19 LTRs of *ras*^{4} and the copia-like transposon

94.9% identity in 409 nt overlap

```
               10        20        30        40        50        60
882    AATTAATTAAATGTATGGTGCAGGTCCCTCGCCGCGGTCTCCGGCGTAGGTTGCAGGTAA
       ::::::::::::::: : ::::::::::::::::::::::::::::::::::::::::::
ras4-5 AATTAATTAAATGTAAGATGCAGGTCCCTCGCCGCGGTCTCCGGCGTAGGTTGCAGGTAA
               10        20        30        40        50        60

               70        80        90       100       110       120
882    CGGGGGGTTCCTCTGTCACTGGGAGGCAGGGCGGTTGCCGCAGACCTCTTCTCCTAGATT
       ::::::::::::::::::::::::::::::::::::: X:::::::::::::::::::::
ras4-5 CGGGGGGTTCCTCTGTCACTGGGAGGCAGGGCGGTTG-CGCAGACCTCTTCTCCTAGATT
               70        80        90        100       110

              130       140       150       160       170       180
882    GGGAGATATGGTGGGAGAACGTCTCTCCGTTGTTGACTGCCCTTAAGGCTAGCCAACCAA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
ras4-5 GGGAGATATGGTGGGAGAACGTCTCTCCGTTGTTGACTGCCCTTAAGGCTAGCCAACCAA
     120       130       140       150       160       170

              190       200       210       220       230       240
882    TTCAATGATAACAGGCAGTTAGCTGGAGTTAGATTGAAGGCGCGATGCGCTCTTTTATTG
       :::::::::::::^v::: ::::::::::: :::::::::::::::::: :: :::::: 
ras4-5 TTCAATGATAACA-GCAG-TAGCTGGAGTT-GATTGAAGGCGCGATGCG-TC-TTTATT-
     180       190         200        210       220         230

               250       260        270         280        290
882    GAATAC-AAATCAAACTGACTATAAGCT-ACAA--GGGAAAACATCATAGCGG-CCTCTG
       :::::: ::::::::::::::::::::: ::::  :: ::::::::::::::: ::::::
ras4-5 GAATACAAAATCAAACTGACTATAAGCTAACAAAGGGAAAAACATCATAGCGGCCCTCTG
           240       250       260       270       280       290

         300       310       320       330        340       350
882    CCAATGCGCAGAGCTTCTGCCGGCTATGCATGAGCTTC-CGGCCAAATGCTTGGTCAGCA
       :::::::::::::::::::::::::::::::::::::: :::::::::::::::::::::
ras4-5 CCAATGCGCAGAGCTTCTGCCGGCTATGCATGAGCTTCACGGCCAAATGCTTGGTCAGCA
           300       310       320       330       340       350

         360       370        380       390       400
882    ATTTGACCGGTGCTGGTG-TGCGGACGATCAGTCCGGTTAACTTAGTTA
       :::::::::::::::::: ::  :::::::::::  :::::::::::::
ras4-5 ATTTGACCGGTGCTGGTGTTGACGACGATCAGTC--GTTAACTTAGTTA
           360       370       380         390       400
----------
```

lfasta sequence comparison of the terminal repeats of the *ras*^{4} insert and the *D. melanogaster* copia-like 8.8kb (DROTN882) transposon in the tropomyosin I gene of flightless mutant *Ifm(3)3*, segment 2 (Accession number K03275). Only the 5′ end LTR sequence of the *ras*^{4} insert is used.

Figure 3-20 37bp deletion in *ras*^{3}

5′ upstream sequences of *ras* gene

gaattcggtg tgcagccaca cagatgcgat atttatcgat agtctcggtg -752
Eco RI
agctgtgcac gaagccagcc atgaaacagt taactgtgac acttctatcg -702

ttactatcga tagctggcat tgtaagcaat agtatcgcag tgcagcaata -652

gctttaagcg gaaagaagct aattcgctgc catattttaa agattaaata -602

aagaggaaat ggatattgct cgagattcag gctattcagg ctaatatgta -552
primer shu 1 »»
atgattattg tttaatatgc ttaataaatt gcatagtcga tattgaaagc -502

gccttagtgt ggctgccatg gaaagcgcga ttaaggccaa agcgtatcgt -452

ggtaatattt agtgggttt*c ccggtggtcc agcgagtata tttaagcggc* -402

*cgcacggcgg ccacac**cacttcacgaactcgattcgcgcagcacacgcgg*** -352

***cac**ttggact ccagcggact ggtaagatgg cccaaccatg tgctcgctgg* -302
^ P element insertion site in LF1
cttgccagcc aatcgggccg attcccattg cgaatcaccc tccgatcgat -252

gcaattaacc ttttccattt gatcccaatt gtctggctgc gccgtgtgcc -202

gtg**ca**tcttc ttcactgccg catttgtgta tgtgccccct gctaattgcg -152
potential transcription start site

tctcaaaagc tcgagaggaa tctgtatcca caaaaagaaa tctgcaacag -102
«« primer shu 2

* The sequence is copied from Figure 3-9. The GC rich region is indicated as following: The *Italic* letters represent 63.6% GC and the underlined letters represent 71.4% GC content. The postulated initiator is labeled. The two arrows indicate the consensus -1C and +1A (bold) in a typical initiator. The P-element insertion site in *ras*^{l24} is also included. The two primers are indicated by the underlines. The 37bp bold letters represent the deletion in *ras*^{3}. There is a trinucleotide CAC at both ends of the deletion.

3.3.5 *ras* revertants

Two *ras* revertants were found in the *ras* mutant stocks by Effie Woloshyn, University of Alberta, during routine stock maintenance. One was a *ras*^{J24} revertant (*ras*^{J24R}), the other a *ras*^{2} revertant (*ras*^{2R}). When first observed as single flies, both of the revertants were heterozygous with unreverted mutant chromosomes. Stocks were established carrying the revertant X chromosomes, both of which gave phenotypically *ras*$^{+}$ homozygous flies. Southern analysis showed that both revertants had lost their inserts (Figure 3-21). Using the four available primers, Primer1/Primer2 for *ras*^{J24R} and Primer3/Primer4 for *ras*^{2R}, the mutant regions were PCR amplified. Direct sequencing of the PCR products showed that excision of the inserts was precise, leaving a perfect wild-type sequence in both revertants. Mutant markers present on the revertant chromosomes in both cases indicated that the appearance of a *ras*$^{+}$ allele was not a result of stock contamination, although, in the case of the reversion of the P-induced lethal mutant, conversion from the wild-type allele carried in the FM6 balancer remains a possibility.

3.3.6 Expression pattern of the *ras* gene

3.3.6.1 Enzyme activity in the late stages of wild-type development

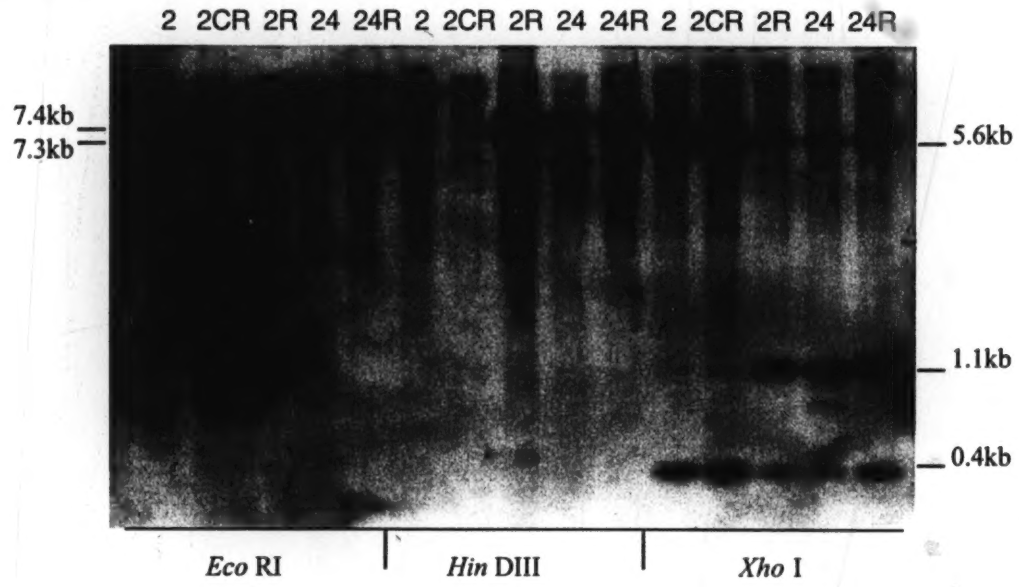
Drosophila eye-colour pigmentation occurs during the end of the second day and early on the third day of the pupal stages. The brownish eye-colour emerges between 3.0-3.5 days of pupation. The greater part of eye-coloration is established well before eclosion. The *ras* strains used in this study show mutant eye-colour which results from defective eye pigmentation and all of these mutants contain a DNA alteration in the IMPDH gene. To further investigate the relationship between IMPDH activity and the eye-colour pigmentation, IMPDH activities in late larvae and in pupae were measured.

The crude extracts were obtained from the same panel of samples as was used in the total RNA extractions. In the protein quantification, a standard BSA concentration curve was established from the average data of two parallel experiments. The protein concentration of each sample was also determined by the average data of two parallel experiments (Appendix 1).

IMPDH activity was identified by observing the production of ^{14}C-XMP derived from ^{14}C-IMP and measured as the amount of ^{14}C-XMP produced per μg protein in crude extracts of whole organisms. In preliminary experiments, samples were collected at 15 minute intervals up to 240 minutes. The results showed that the reaction ceased after 30 minutes. When samples were taken at 6 minute intervals, the result showed that the reaction curve was nearly linear in the first 12 minutes and slowed down after that (Figure 2-4). Based on these results, three time points (3, 6 and 9 minutes) were used in subsequent assays.

IMPDH activity was examined from the late larval to the young adult stage, which includes the period when eye-pigmentation is developing. The highest level of activity occurs in the first day and the fourth day pupal stages, with relatively low levels in larval, second and third day pupal, and young adult stages (Figure 3-22, 3-23). There is no obvious difference between males and females.

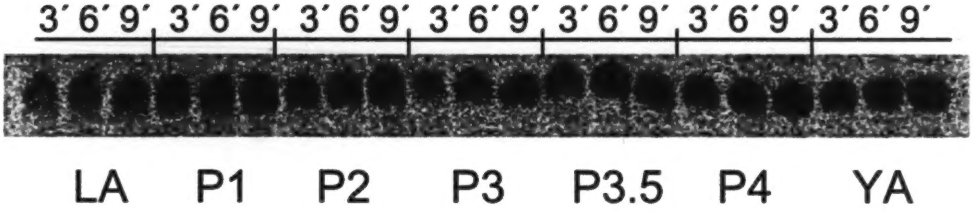
Figure 3-21 Southern analysis of *ras* revertants



Hybridization of ^{32}P-labeled *Eco* RI-*Hin* dIII DNA fragment to Southern-transferred restriction digests of genomic DNA from the revertants of *ras*^{2} (*ras*^{2R}) and *ras*^{l24} (*ras*^{l24R}) females. The size markers are labeled the same as in Figure 3-12 and Figure 3-13. The original *ras*^{2R} was heterozygous for *ras*^{2R}/*ras*^{2}. Both the mutant chromosome and the revertant chromosome in the *ras*^{2R} revertant were isolated. The mutant chromosome was used in this experiment as a control. The samples of *ras*^{2} and the mutant chromosome of *ras*^{2R} were overloaded so that they run faster than the size markers. Both homozygous *ras*^{2R} and *ras*^{l24R} females show wild-type digestion pattern.

2 = *ras*^{2} (homozygous)
2CR = *ras*^{2R} (derived from the mutant chromosome of the revertant, homozygous)
2R = *ras*^{2R} (derived from the wild-type chromosome of the revertant, homozygous)
24 = *ras*^{l24}/*FM6*
24R = *ras*^{l24R} (derived from the revertant chromosome of *ras*^{l24}, homozygous)

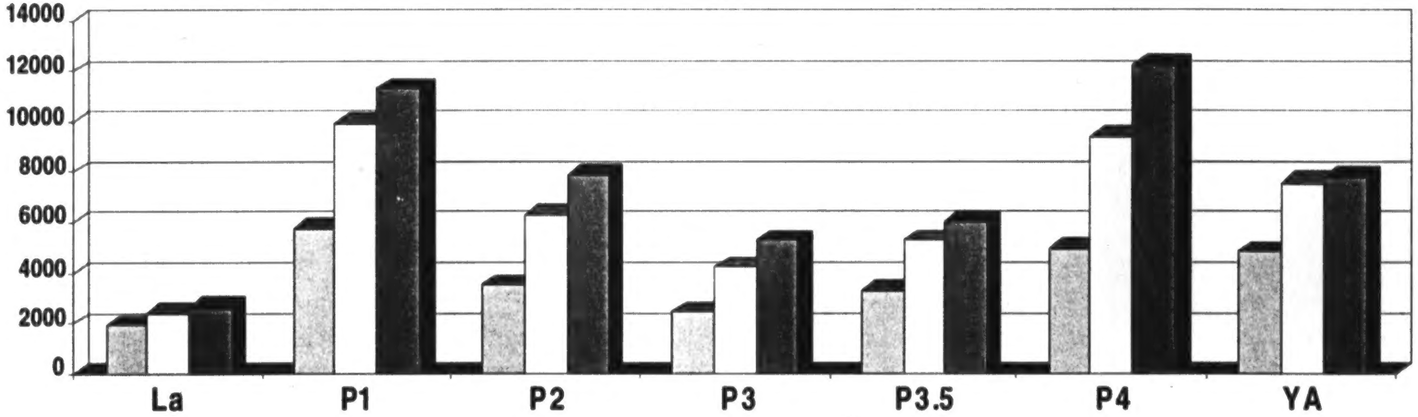
Figure 3-22 IMPDH activity of the wild-type females during later stages of development



PhosphorImager scan of the chromatographically separated IMPDH product (^{14}C-XMP) from crude protein extracted from wild-type females

^{14}C-XMP was converted from ^{14}C-IMP by IMPDH in the crude protein extracts from wild-type females followed by chromatography (*Section 2.6*). The figure only shows the ^{14}C-XMP signal on the TLC plate. Samples were taken at 3 (3′), 6 (6′) and 9 (9′) minutes of the reaction.

Figure 3-23 IMPDH activity of wild-type females during later stages of development



This figure shows the ^{14}C-XMP single adjusted by the crude protein concentration. The data were generated from the Figure 3-22 and the detailed calculation is in Appendix 4.

3.3.6.2 IMPDH activities in late developmental stages of *ras* mutants

IMPDH activity was also examined in *ras*$^{1,\ 2\ and\ 3}$ from larval to young adult stages (Figure 3-24, 3-25). In *ras*^{3}, because the mutant eye-colour shows solely in males, IMPDH activity was examined in both sexes. In *ras*^{1} and *ras*^{2}, only females from both strains were examined. IMPDH activity was only examined in *ras*^{4} young adult females because of the infertility of *ras*^{4}/*ras*^{4} females. To make a direct comparison between different strains, samples were arranged according to developmental stages.

The major difference observed between wild-type and mutants is in the fourth day pupae and young adult stages. The activity in *ras*$^{1,\ 2\ and\ 3}$ is much lower than that in wild-type in these two stages (Figure 3-24). The low IMPDH activity in young adult females of these mutants was also found in preliminary experiments, but the difference is not so dramatic (Figure 3-26). Unlike in the above *ras* mutants, the activity in *ras*^{4} is higher than the wild-type. The higher *ras*^{4} IMPDH activity was observed constantly in three experiments conducted so far (Figure 3-24, 3-26). Another observation is the difference between males and females in *ras*^{3}. The IMPDH activity in males is about half of that in females in the third and fourth day pupal stage. This observation agrees with the lower IMPDH RNA level of *ras*^{3} males during the same developmental stages.

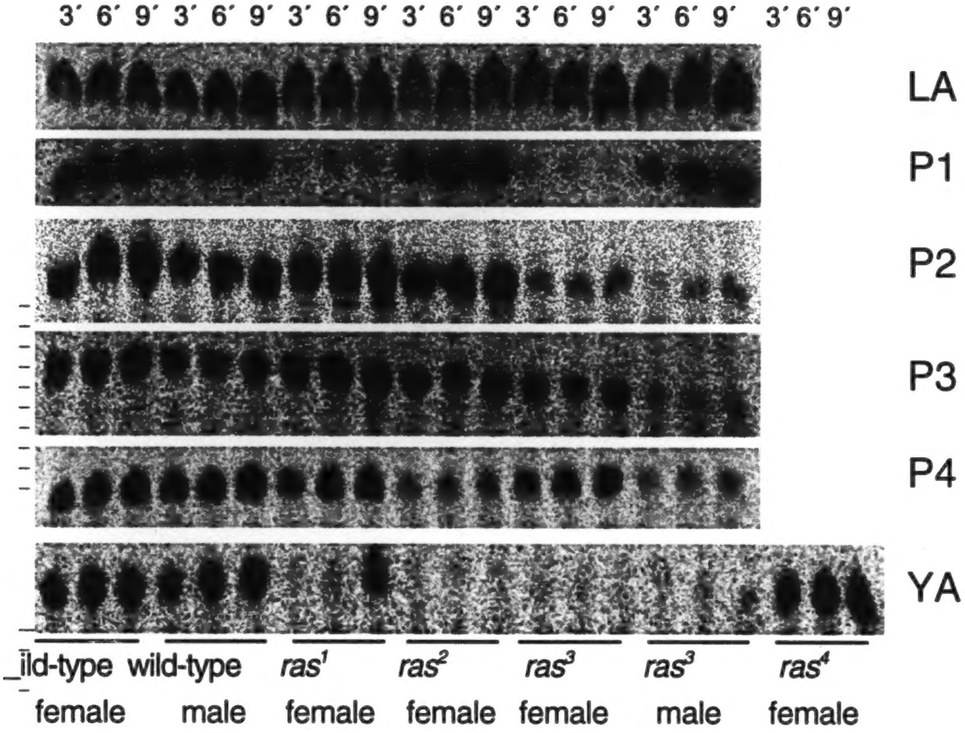
The extremely low IMPDH activity levels of *ras*^{1} and *ras*^{3} females in the first day pupae seems strange. Since no other experiments involving this stage were conducted, the meaning of this low activity is unknown. The high level of the nine minute sample of *ras*^{1} young adult females is clearly inconsistent with the remaining results; it is presumably due to contamination. Examination of the autoradiographic image suggests it to be a tiny dot with an extremely high level of radioactivity (Figure 3-24).

3.3.6.3 Developmental northern analysis of wild-type strains

To study the expression pattern of the *ras* gene, the developmental RNA profiles of the wild-type were examined. Northern analysis of total RNA from different developmental stages was conducted. The RNA probe ("ribo-probe") used in the study was synthesized from the first ~0.8kb of the cDNA KS29 to avoid the "OPA" sequences at the 3′ end of the template, which certainly generates multiple bands during hybridization to genomic DNA (Figure 3-10). A similar phenomenon was apparent in preliminary northern analyses using a probe from the entire gene (data not shown).

The following analysis is based mainly on the northern data generated from a single experiment. To standardize the conditions, all northern blots were hybridized in the same hybridization bag then exposed to X-ray films for the same time. Several other experiments were performed with RNA samples extracted separately to confirm the major differences between wild-type and *ras* mutants, especially during the early developmental stage (0-16 hours). In an attempt to reinforce the quantification accuracy, a set of hybridized filters was quantified by the phosphorimaging method. Unfortunately, the phosphorimaging results were very incomplete (due to heavy background and patchy transfer) and failed to yield a complete RNA profile analysis. They did confirm qualitatively most of the major hybridization results from the autoradiographic analysis.

Figure 3-24 IMPDH activity in late developmental stages

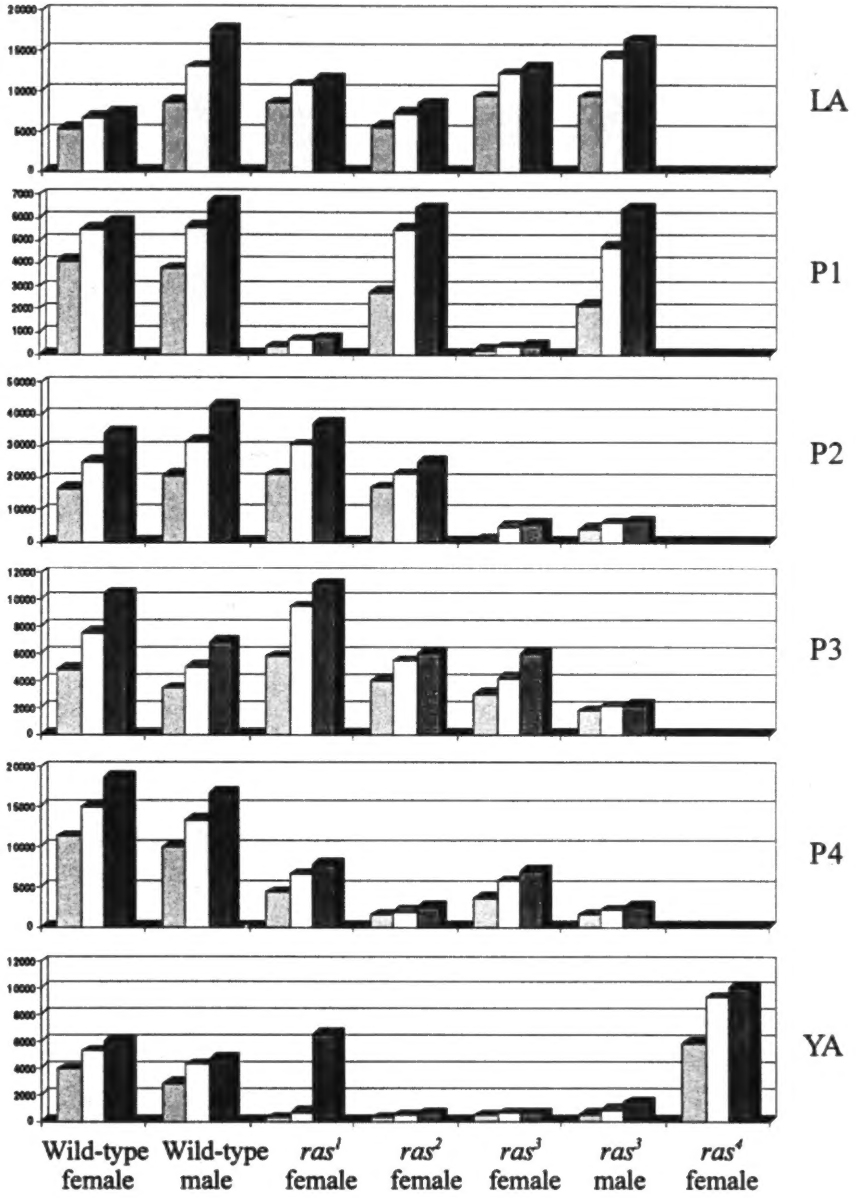


PhosphorImager scan of the chromatographically separated IMPDH product (^{14}C-XMP) by crude protein extracted from different strains

The analysis was performed similarly to that shown in the Figure 3-24. The material was collected from wild-type and *ras* mutants during pupal development. Samples were arranged by the developmental stages (labeled at the left side of the corresponding rows). The samples were taken at 3 minute intervals. The signal at 9′ of young adult stage of *ras*^{1} was due to contamination.

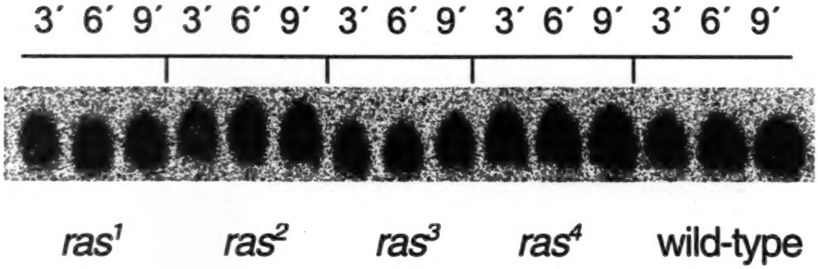
3′	3 minute sample
6′	6 minute sample
9′	9 minute sample

Figure 3-25 IMPDH activity in late developmental stages



This figure shows the ^{14}C-XMP single adjusted by the crude protein concentration from different strains during development. The data were generated from the Figure 3-27 and the detailed calculation is in Appendix 5.

Figure 3-26 IMPDH activities of young adult females



PhosphorImager scan of the chromatographically separated IMPDH product (^{14}C-XMP) by crude protein extracted from young adult females

The analysis was performed similarly to that shown in the Figure 3-25. The ^{14}C-XMP yields in *ras*^{1}, *ras*^{2} and *ras*^{3} are lower than that in wild-type and the activity level in *ras*^{4} is higher than that in wild-type.

in the figure,

1	*ras*^{1}
2	*ras*^{2}
3	*ras*^{3}
4	*ras*^{4}
+	wild-type
3′	3 minute sample
6′	6 minute sample
9′	9 minute sample

During the analysis, it was noticed that sometimes the density of the signals was not constant between different experiments. This inaccuracy made the interpretation of the data difficult. The inconsistency mostly occurs in the early developmental stages and with weak signals. The difference may be caused by experimental errors such as the time overlapping in the collections of samples of 0-2, 2-4 and 4-8 hour embryos. To avoid this kind of inaccuracy, several experiments were repeated several times to identify the most likely RNA profile.

Relative to RP49 loads (and also, probably, in absolute terms), the highest IMPDH RNA level in wild-type is found in the 0-2 hour embryos. It is at least 3-4 times the level found in young adults. The level rapidly drops down to the lowest level at 2-4 hours, when it is less than one tenth of the young adult level. Transcript level starts to increase again at 4-8 hours, suggesting the zygotic transcription occurs at this stage. The level peaks at 8-16 hours, before it drops again and stays relatively low during the remainder of embryonic and most of larval life. At the later third instar stage, the transcript level increases to a little higher than that of young adults and stays high in most of the pupal stage, and only temporarily drops in samples taken from the 3.5 and 4 days in the pupal life (Figure 3-27, 3-28, 3-29; Appendix 4, 5).

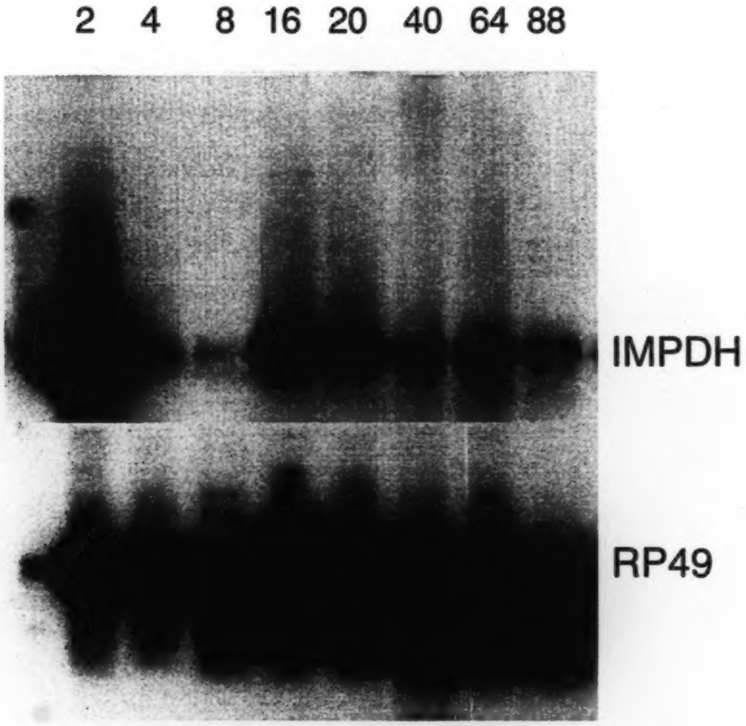
Some apparent sex related differences of RNA level were identified in this experiment. In late larval and early pupal stages, the RNA level in males is almost two times the level in females. On the third and fourth day of the pupal stage, the RNA level of females is higher than that of males (Figure 3-28). These differences can be observed in Figure 3-28 by comparing the RP49 and the IMPDH signals. A similar sex related difference has not been observed in other experiments. When considering the data generated from all experiments, there seems no obvious sex related difference during development.

3.3.6.4 The transcript pattern in the late developmental stages of *ras* mutants

To study how the *ras* mutant lesions affect *ras* gene expression, RNA expression patterns of the four *ras* viable mutants were also investigated. There is an additional ~7kb band in each of the *ras*$^{1,\ 2\ and\ 4}$ total RNA samples from young adults (Figure 3-30, 3-31, Figure 3-32, 3-33, Figure 3-34, 3-35). The size of the extra RNA is approximately the combined size of IMPDH mRNA (~2.4kb) and the individual inserts in these mutants, which are about 4-5kb. Based on the size information, these extra RNA molecules are possibly the IMPDH transcripts with the unspliced second intron and the nested inserts. The extra band was not observed in a preliminary northern analysis of *ras*^{2} Poly-A$^+$ RNA (Figure 3-6), implying it is not polyadenylated.

In *ras*^{1} (Figure 3-30) the large band is also present in the late larval to the young adult stages with a constant size. The weak signal of this large band in the late larval female samples may due to bad transfer or hybridization. In *ras*^{2} (Figure 3-32), the larger band is clearly present in the first and fourth day female pupae and in the third day pupae and young adult of both sexes but it is not (or barely) observed in the other samples. The size of the unspliced transcript also shows some differences: In the first and third day pupae, the female unspliced band is a little larger than the unspliced band in other samples, including the third day male unspliced transcripts. The reason for the size difference is unknown.

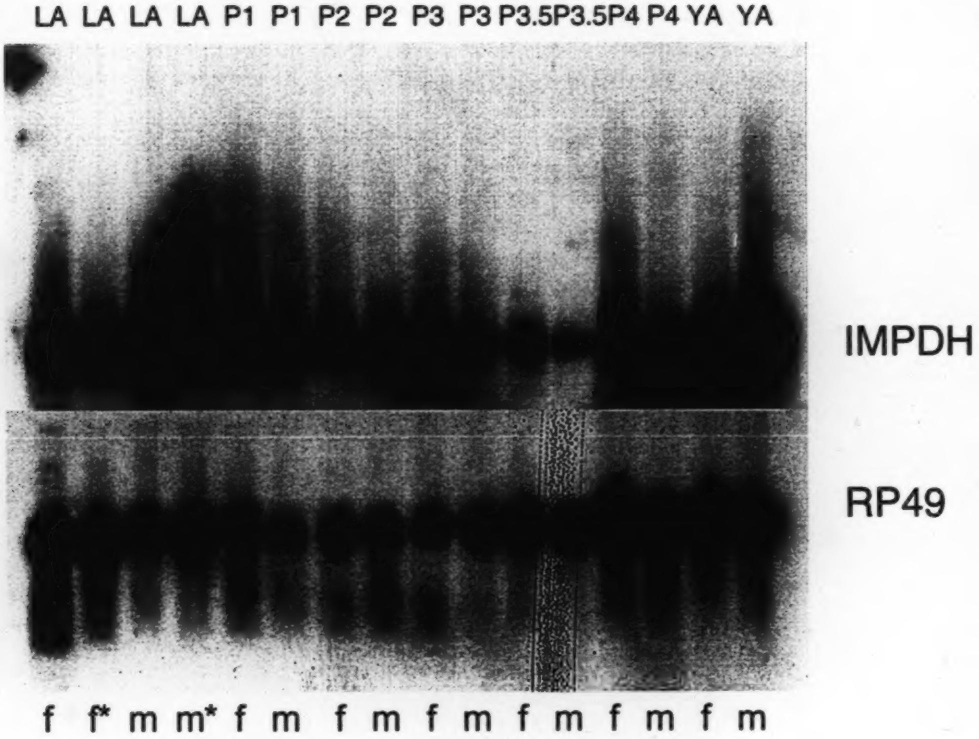
Figure 3-27 Northern analysis of 0-88 hours wild-type RNA



The northern blot was hybridized with the IMPDH probe (the first 800bp of cDNA 29) first. After exposure the filter was stripped, then re-hybridized to the RP49 probe. To show the comparison between the IMPDH and RP49 signals, the two films were aligned according to the original relative positions on the films. The data in Table 3-4 were generated from these images.

2 = 0-2 hours embryos
4 = 2-4 hours embryos
8 = 4-8 hours embryos
16 = 8-16 hours embryos
20 = 16-20 hours larvae
40 = 20-40 hours larvae
64 = 40-64 hours larvae
88 = 64-88 hours larvae

Figure 3-28 Northern analysis of wild-type RNA from late larval to young adult stage


LA LA LA LA P1 P1 P2 P2 P3 P3 P3.5 P3.5 P4 P4 YA YA
IMPDH
RP49
f f* m m* f m f m f m f m f m f m

The analysis was performed similarly to that shown in the Figure 3-22. Unless otherwise noted, the material was collected from Am$^+$, an inbred derivative of Oregon R.

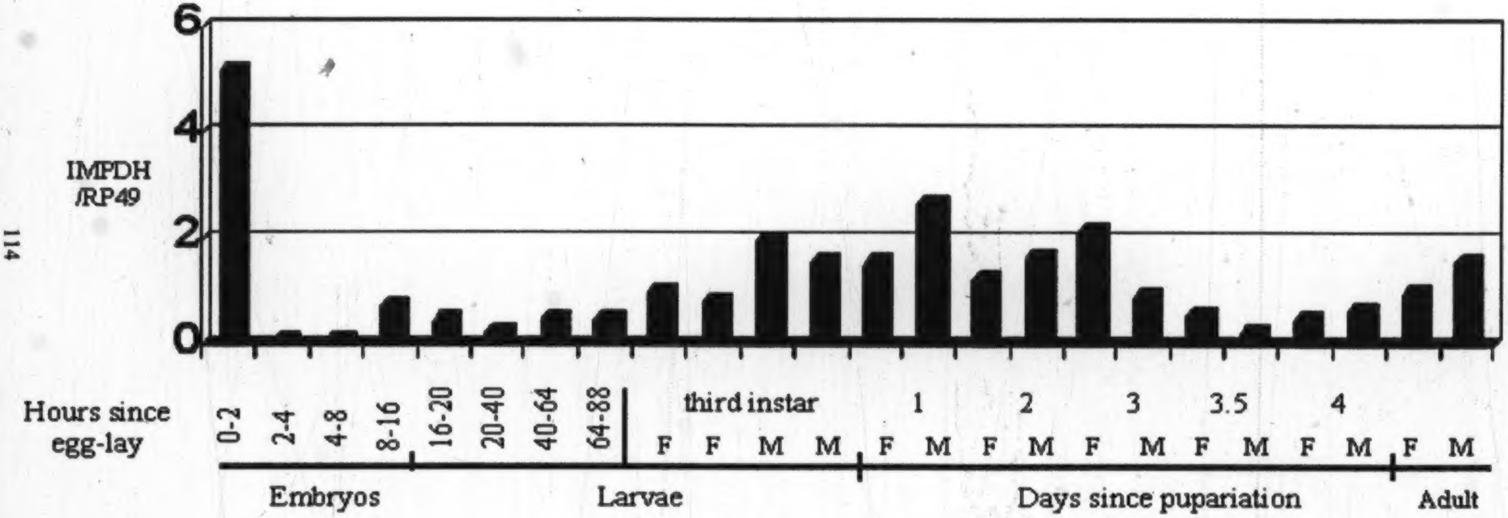
In the figure,

LA = late larval stage (third day instar)
P1 = pupae 1, one day after sexing late third instar larvae
P2 = pupae 2, two days after sexing late third instar larvae
P3 = pupae 3, three days after sexing late third instar larvae
P3.5 = pupae, 3.5 days after sexing late third instar larvae
P4 = pupae, four days after sexing late third instar larvae
YA = young adult (0-8 hours)

f = female
m = male

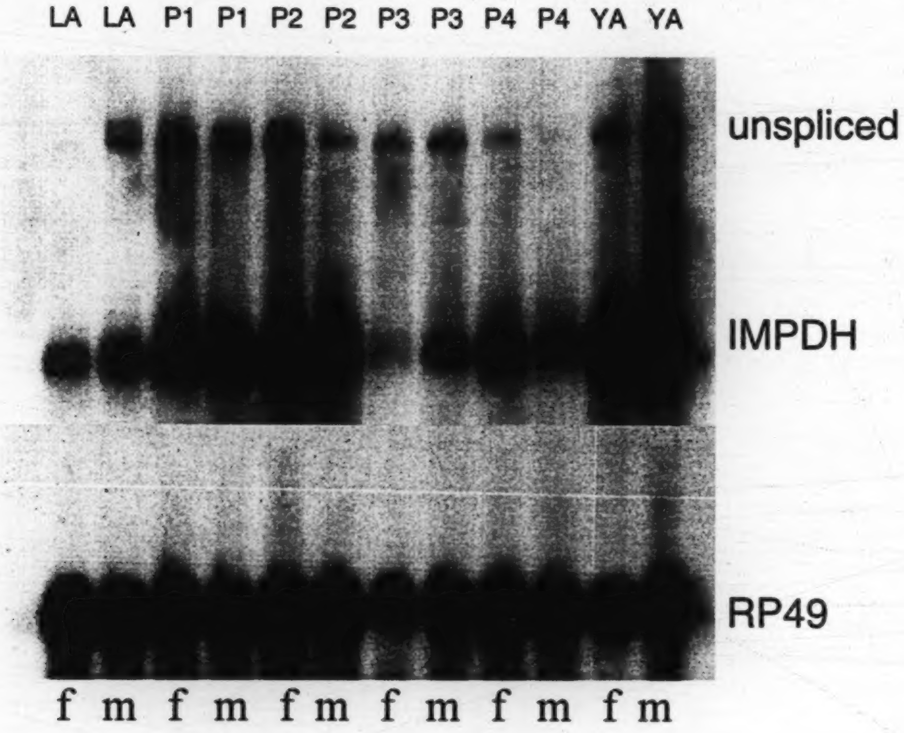
*The same legends are used in all the following figures.

Figure 3-29 Wild-type IMPDH RNA levels in 0-88 hour Drosophila embryos



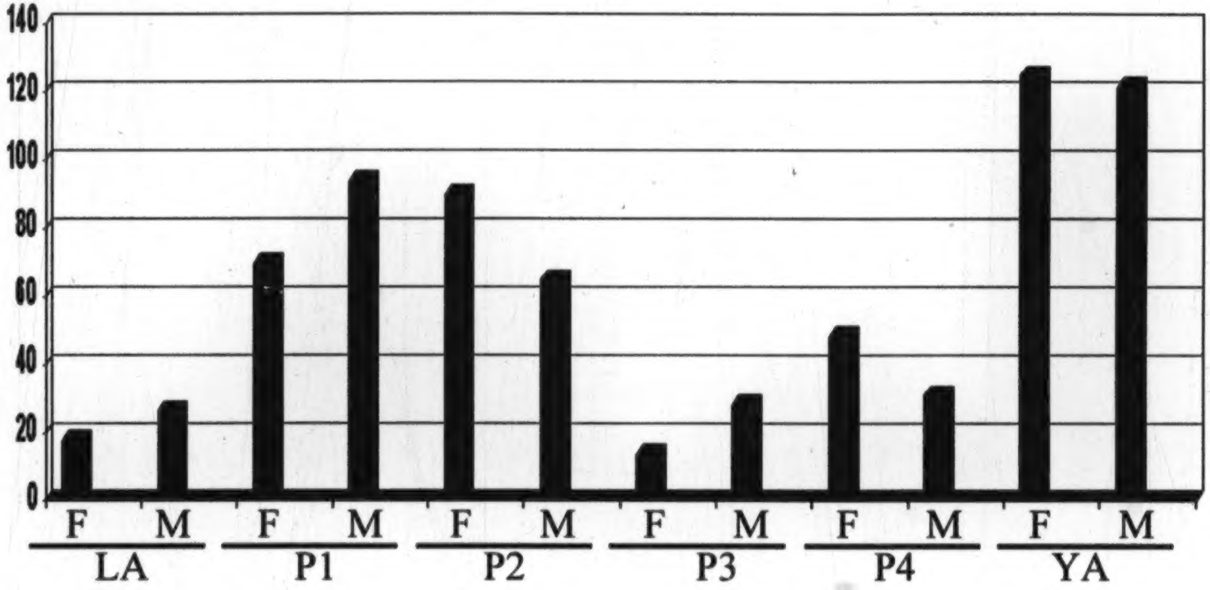
This figure shows the wild-type developmental profile of the relative IMPDH transcript levels adjusted by RP49 transcript in Drosophila. The data were generated from Figure 3-22 and 3-23 and the detailed calculation is in Appendix 1 and 2.

Figure 3-30 Northern analysis of *ras*^{1} late larvae to young adult samples



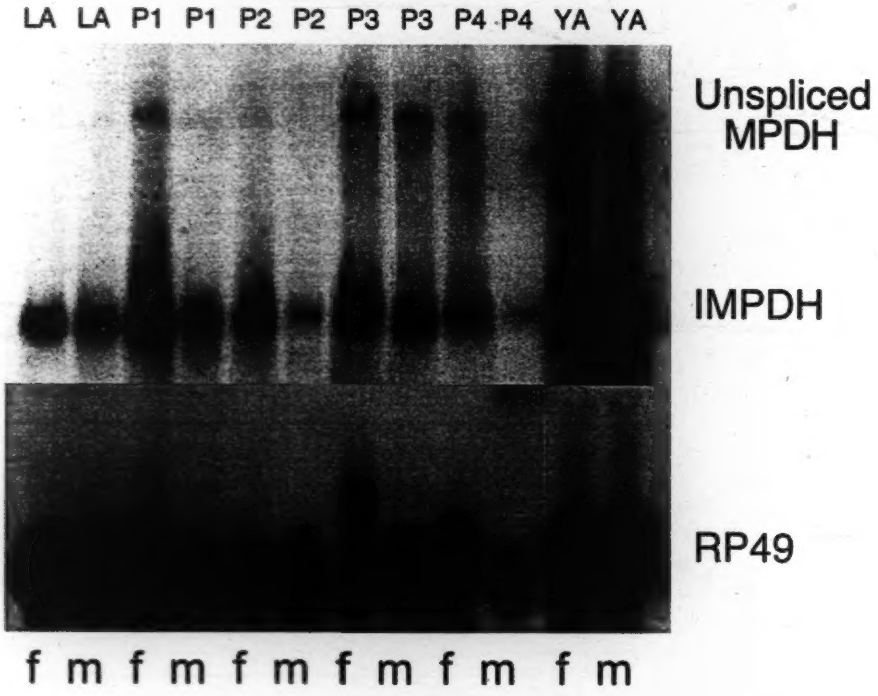
The analysis was performed similarly to that shown in the Figure 3-22.

Figure 3-31 Northern analysis of *ras¹* late larvae to young adult samples



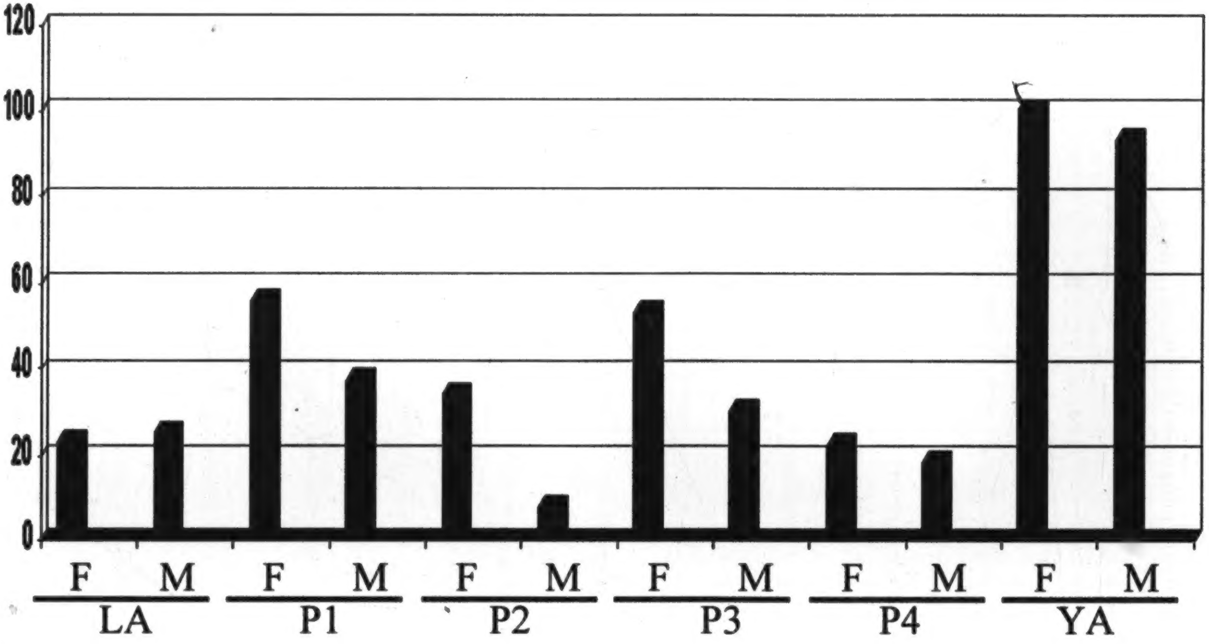
This figure shows the developmental profile of the relative IMPDH transcript levels adjusted by RP49 transcript in *ras¹*. The data were generated from Figure 3-30 and the detailed calculation is in Appendix 6.

Figure 3-32 Northern analysis of *ras*^{2} late larvae to young adult samples



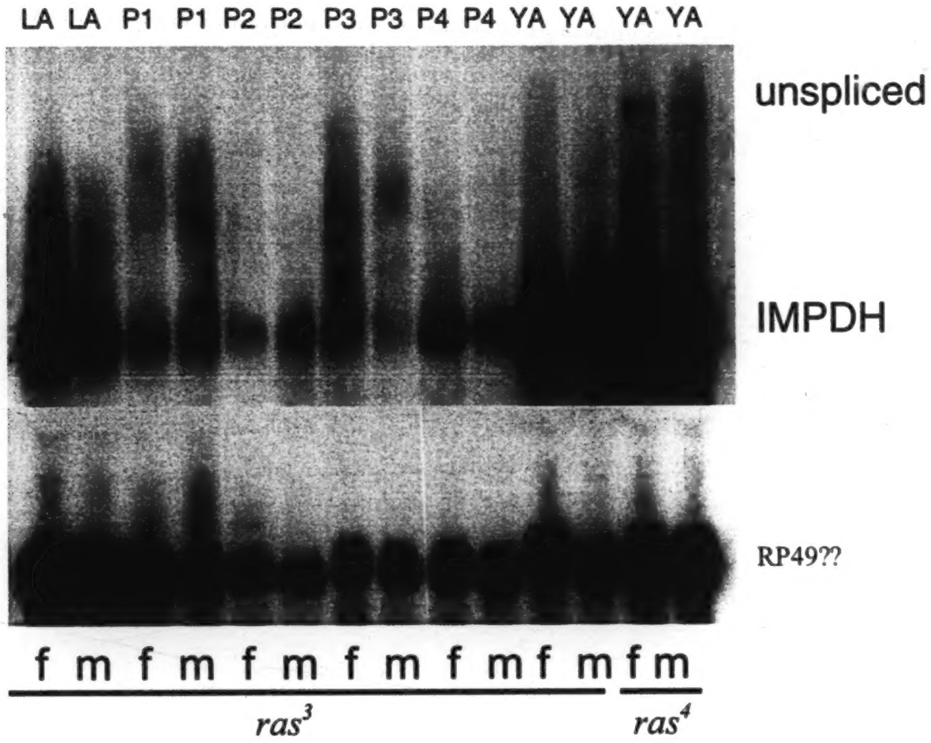
The analysis was performed similarly to that shown in the Figure 3-22.

Figure 3-33 Northern analysis of *ras*^{2} late larvae to young adult samples



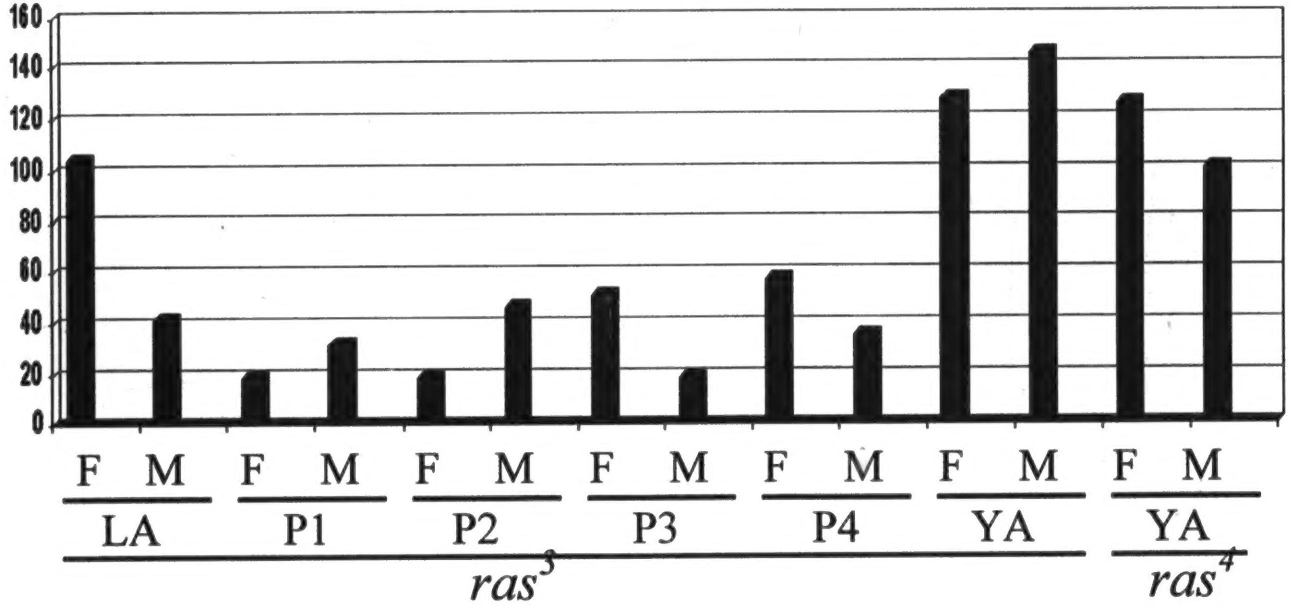
This figure shows the developmental profile of the relative IMPDH transcript levels adjusted by RP49 transcript in *ras*^{2}. The data were generated from Figure 3-32 and the detailed calculation is in Appendix 7.

Figure 3-34 Northern analysis of *ras*^{3} late larvae to young adult and *ras*^{4} young adult samples


LA LA P1 P1 P2 P2 P3 P3 P4 P4 YA YA YA YA
unspliced
IMPDH
RP49??
f m f m f m f m f m f m f m
ras3
ras4

The analysis was performed similarly to that shown in the Figure 3-22. Unless otherwise noted, the material was collected from *ras*^{3}.

Figure 3-35 Northern analysis of *ras³m* late larvae to young adult and *ras⁴* young adult samples



This figure shows the developmental profile of the relative IMPDH transcript levels adjusted by RP49 transcript in *ras³m* and *ras⁴*. The data were generated from Figure 3-34 and the detailed calculation is in Appendix 8.

The IMPDH RNA level in *ras*$^{1 \text{ and } 2}$ is lower than in the wild-type during the late larval and the pupal stages (Figure 3-30, Figure 3-32). In the young adult, the transcript level increases to the normal range in both mutants. The transcript level in *ras*^{4} young adult is also in the wild-type range (Figure 3-34).

Some differences between the sexed samples were observed. In *ras*^{1}, the signals in the third day females are much lower than that in males (Figure 3-30). Since both smaller (RP49) and larger (unspliced transcript) bands are clearly present in the same lane and their densities are comparable to the same bands in other lanes, the low IMPDH RNA signal is unlikely caused by technical errors. In *ras*^{2}, the IMPDH signals in the second and fourth day males are lower than that in females (Figure 3-32). However, since the signals of the unspliced bands are very weak in both cases, it is not clear if the difference is caused by technical errors. No constant sexual differential pattern was observed in other experiments, so the meaning of the difference is uncertain.

The RNA level of *ras*^{3} is low in the later larval and the whole pupal stages compared to the wild-type. This is similar to the pattern in both *ras*^{1} and *ras*^{2} mutants (Figure 3-34). The RNA level in males is low in the late larvae, P3 and P4 stage compared to that in females. The sexual difference in the early stage (late larval, the first day and second day pupal) was not observed in the phosphorimaging experiment.

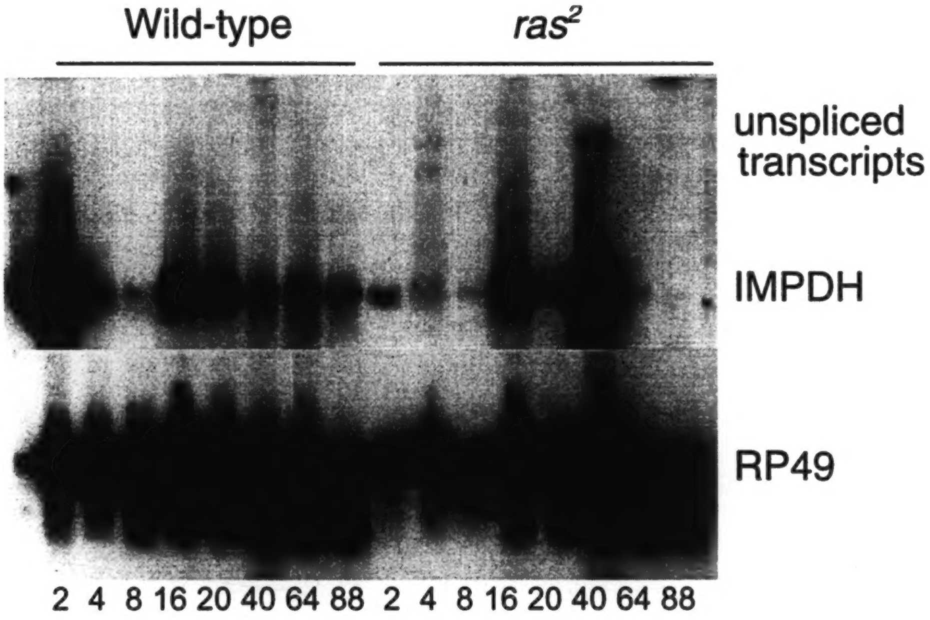
3.3.6.5 The transcript pattern in the early developmental stages of *ras* mutants

To study the early developmental pattern of IMPDH transcripts in the *ras* mutants, *ras*^{2} embryos were chosen to conduct northern analysis along with wild-type embryos, since *ras*^{2} is the most examined *ras* mutant in our laboratory (Figure 3-36, 3-37). The experiment reveals that there are almost no transcripts in the *ras*^{2} early embryo stage. The RNA level of 0-2 hours is strikingly low, given that the wild-type exhibits the highest level in this same period. This is a surprising result. To further investigate early IMPDH RNA levels, northern analysis was performed with *ras*$^{1,\, 2 \text{ and } 3}$ embryos. This time, the lack of IMPDH RNA was clearly showed in *ras*^{2} 0-2 hour embryos, as well as in another insertion mutant, *ras*^{1}, during the same developmental stage (Figure 3-38). The 16-88 hour RNA samples were also examined. Unfortunately, because of bad gel electrophoresis and hybridization, no complete data were collected.

Since there is little zygotic RNA synthesized at this stage in Drosophila, lack of RNA in the 0-2 hour stage can be explained by the lack of maternal transcripts. However, there is another possibility. If the *ras*^{1} or *ras*^{2} females hold the fertilized eggs two hours longer than the wild-type females because of the mutation, the newly laid eggs would lack RNA since the maternal transcripts would have been degraded and the zygotic RNA has not been synthesized. To test this hypothesis, the developmental morphological change of the newly laid eggs was inspected. There are no obviously morphological differences observed between the wild-type and *ras*$^{1,\, 2 \text{ and } 3}$ eggs up to 150 minutes after the egg lay (Table 3-5). So, the lack of IMPDH RNA in the 0-2 hour embryos in *ras*^{1} and *ras*^{2} seems due to the lack of maternal transcripts.

Further analysis of Figure 3-38 has revealed that the wild-type zygotic transcripts appear at about 4-8 hours. There are almost no transcripts at the same period in either *ras*^{1} or *ras*^{2}. This indicates that zygotic transcription is also delayed or reduced in both *ras* insertional mutants.

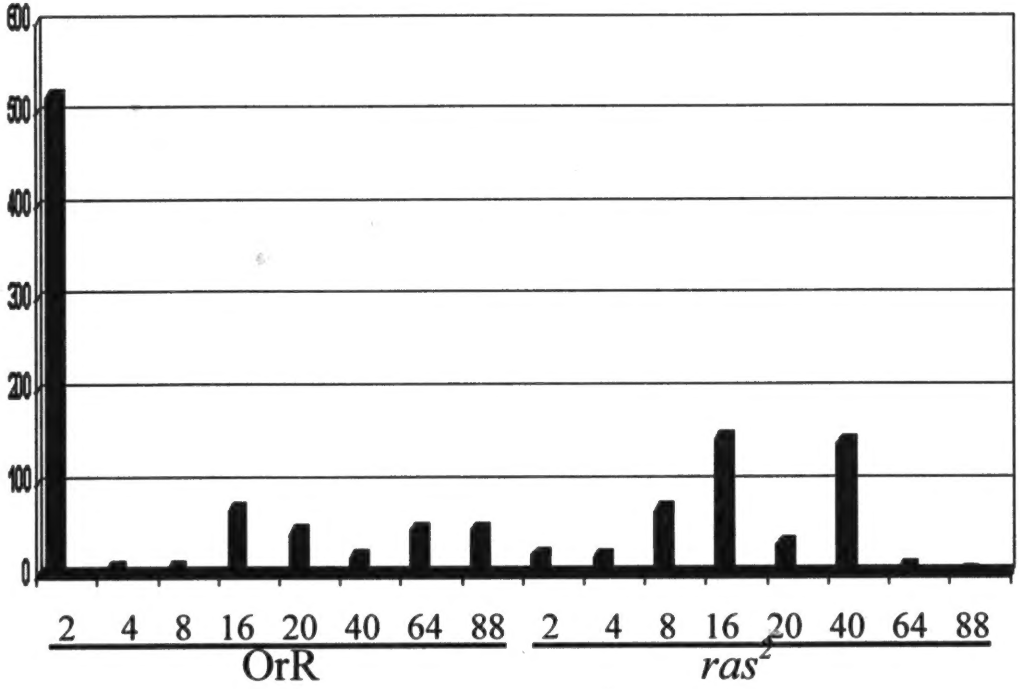
Figure 3-36 Northern analysis of 0-88 hour Am^{+} and *ras*^{2} samples



The northern blot was hybridized with the IMPDH probe first. After exposure the filter was stripped, then re-hybridized to the RP49 probe. To show the comparison between the IMPDH and RP49 signals, the two films were aligned according to the original relative positions on the films.

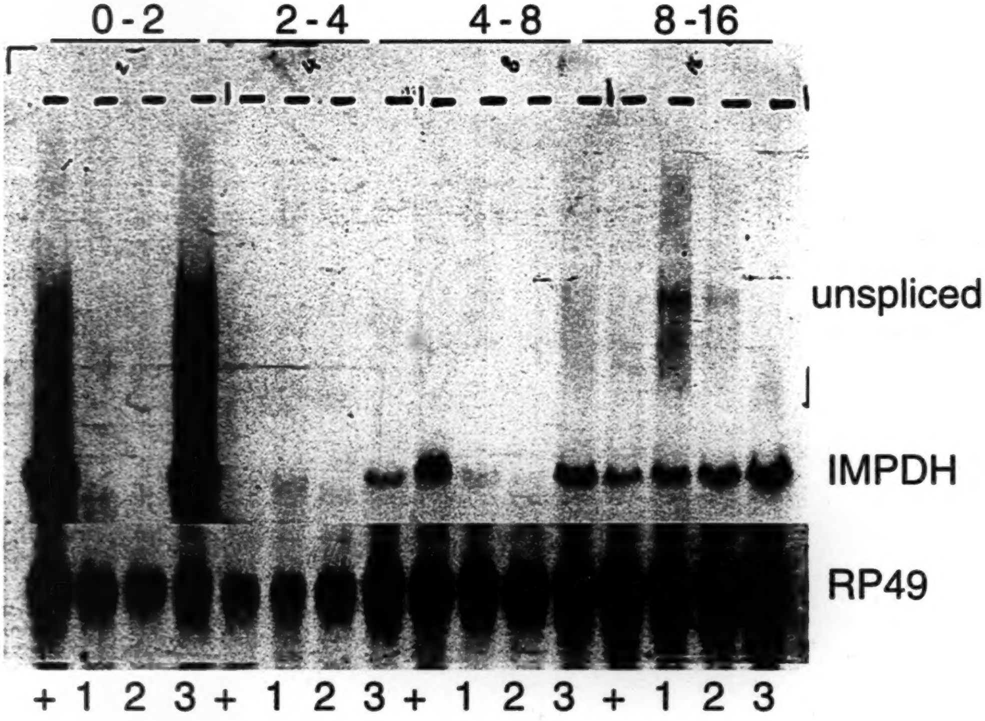
2 = 0-2 hours
4 = 2-4 hours
8 = 4-8 hours
16 = 8-16 hours
20 = 16-20 hours
40 = 20-40 hours
64 = 40-64 hours
88 = 64-88 hours

Figure 3-37 Northern analysis of 0-88 hour Am^{+} and *ras*^{2} samples



This figure shows the profile of the relative IMPDH transcript levels adjusted by RP49 transcript of wild-type and *ras*^{2} 0-88 hour samples. The data were generated from Figure 3-36 and the detailed calculation is in Appendix 9.

Figure 3-38 Northern analysis of 0-16 hour samples of OrR and *ras*$^{1-3}$



0-2 = 0-2 hours
2-4 = 2-4 hours
4-8 = 4-8 hours
8-16 = 8-16 hours

+ = OrR
1 = *ras*^{1}
2 = *ras*^{2}
3 = *ras*^{3}

Table 3-5 Developmental morphological change in the early embryos

A.

Time	0 - 30 min.			30 - 60 min.			60 - 90 min.			90 - 120 min.			120 - 150 min.		
Stage	1-3	4-5	≥ 6	1-3	4-5	≥ 6	1-3	4-5	≥ 6	1-3	4-5	≥ 6	1-3	4-5	≥ 6
OrR	9	1	0	9 ? 1	3	0	5	6	1	6	7	1	1	4	6
ras^{2}	8 ? 1	1	0	11	2	0	9 ? 1	3	0	3	8	0	1	11	1

B.

Time	0 - 30 min.			30 - 60 min.			60 - 90 min.			90 - 120 min.			120 - 150 min.		
Stage	1-3	4-5	≥ 6	1-3	4-5	≥ 6	1-3	4-5	≥ 6	1-3	4-5	≥ 6	1-3	4-5	≥ 6
OrR	10			11	2		8	2	1	6	3	1	4	8	1
ras^{2}	12	1 ?		12	1		12	2		7	4	1	3	7	2
ras^{1}	14			13	1		15	1		7	15	2	2	8	4
ras^{3}	12			12	1		16	2		5	7	2	3	6	2

Young females were transferred to small egg collection chambers on a Petri dish with apple juice/agar medium. Freshly laid eggs were collected at 30 minute intervals at room temperature by changing Petri dishes. The collected eggs were maintained at room temperature in the original dishes. The morphological changes of the eggs were checked at 30 minute intervals. The following morphological stages were described by Wieschaus and Nüsslein-Volhard, 1986. The number under each morphological stage represents the number of eggs at that stage. No developmental delay was found in either *ras^{1}* or *ras^{2}* in the first two and one half hours. A. and B. represent two separate experiments.

1. freshly laid egg
2. early cleavage
3. pole cell formation
4. syncytial blastoderm
5. cell formation
6. early gastrulation

? uncertainty

The IMPDH RNA profile of *ras*^{3} is similar to the RNA profile of wild-type in the first 16 hours. It is high at 0-2 hours and rapidly drops to almost the lowest level at 2-4 hours. The zygotic transcript appears at four hours (Figure 3-38).

In general, the highest level of IMPDH RNA was found in the 0-2 hour embryos in the wild-type strain, which is presumed to represent inherited maternal IMPDH transcripts. The zygotic IMPDH transcription starts at about 4-8 hours of the embryonic stage. In the insertion mutants, the transcript levels and enzyme activities of *ras*^{1} and *ras*^{2} are lower than that of wild-type in late developmental stage. However, the enzyme activity in *ras*^{4} young adult, which is the only stage examined, is much higher than the wild-type activity. In the deletion mutant *ras*^{3}, the levels of RNA and enzyme activity are also lower than the wild-type. There may be some sex related differences of both the RNA level and enzyme activity in *ras*^{3}: The levels in males are lower than those in females. In addition, both *ras*^{1} and *ras*^{2} lack maternal IMPDH transcripts. In the wild-type flies, the highest level of IMPDH transcripts occurs in the 0-2 hour embryos, implying a high level of IMPDH transcripts is required for embryogensis. However, both *ras*^{1} and *ras*^{2} are fully viable and both have almost no IMPDH transcripts in 0-2 hour embryos. Thus, the function and the amount of the maternal IMPDH transcripts seem contradictory in the early embryos. The zygotic IMPDH transcription in both *ras*^{1} and *ras*^{2} is also delayed (occurs between 8-16 hours). The similar behavior of these two mutants may due to the similar DNA alterations found in both.

4 Discussion

4.1 IMPDH and the *ras* complex

4.1.1 The 4.0kb *Eco* RI-*Hin* dIII fragment and the *ras* locus

This thesis focuses on the molecular characterization of the *Drosophila melanogaster raspberry (ras)* locus. It involved cloning the *ras*$^+$ gene, identifying its molecular structure and function, pinpointing the molecular lesions of the four spontaneous *ras* eye-colour viable mutants, and studying the expression of the *ras* gene in both wild-type and *ras* eye-colour mutants.

Cloning the *ras*$^+$ gene was the foundation of the thesis. By using a DNA probe obtained by P-element tagging of the *ras* gene (Leonard, 1986), a λ recombinant phage (SH4) was isolated from a wild-type Am$^+$ (P-element free) genomic library. *In situ* hybridization confirmed that SH4 is located at the *ras* locus. Study of a group of four spontaneous *ras* eye-colour mutants, *ras*$^{1\text{-}4}$, and five P-element induced *ras* recessive lethal mutants, *ras*$^{l22\text{-}l26}$, revealed that all of them show DNA alterations in a 4.0kb *Eco* RI-*Hin* dIII fragment of SH4 (*Section 3.3.1*).

Analysis of spontaneous revertants from two insertional mutations, *ras*^{l24R} (from *ras*^{l24}) and *ras*^{2R} (from *ras*^{2}), shows that both inserts have been completely excised from the 4.0kb region and the insertion sites in both cases have been precisely restored to the wild-type sequence (*Section 3.3.5*). This evidence strongly suggests that, in the two *ras* mutants, DNA alterations in the 4.0kb *Eco* RI-*Hin* dIII fragment are the most likely cause of the mutant phenotype and the 4.0kb region probably is the location of the *ras* gene. It suggests, as well, that the mutant phenotypes in all nine *ras* mutants result from the sequence alterations identified.

Sequence analysis indicates that the 4.0kb *Eco* RI-*Hin* dIII fragment contains an entire translational unit (including four essentially conventional introns) which encodes an IMPDH-like amino-acid sequence. Thus the 4.0kb region is also the location of the Drosophila IMPDH gene. The presumed transcriptional initiation site is also located in this region. There is not yet evidence to prove that the IMPDH gene is restricted to the 4.0kb fragment. It is quite possible that some transcriptional regulatory sequences are present in the 5′ and/or the 3′ untranscribed regions outside the 4.0kb fragment.

Much of this work, along with the genomic DNA sequence and cDNA sequence, has been reported in abstract by Nash and Hu (1992) and in full by Nash *et al.* (1994). Sifri *et al.* (1994) reported a 2.4kb cDNA encoding an entire IMPDH protein, isolated from a Drosophila head library with a mouse IMPDH probe. *In situ* hybridization located the cDNA at 9E1-4 on the X chromosome. The authors noted that 9E1-4 is the locus of the *ras*-complex, which had previously been reported to be involved in the synthesis of GMP from IMP (Falk and Nash, 1974b).

Slee and Bownes (1993) also reported, in abstract, the existence of a similar Drosophila IMPDH cDNA associated with an anonymous gene located in the 8C-9E region. This cDNA was identified in an enhancer-trap screen for testis-specific expression. Northern analysis revealed that several distinct transcripts are produced from the gene. Study of three mutants with mitotic defects in the 8C region, *discs degenerate 3, 9 and 11 (l(1)dd3, 9 and 11, or d.deg-3, 9 and 11),* showed that no abnormalities that could be detected by the IMPDH coding sequence. Later, in a similar Southern analysis of the four *ras* viable mutants conducted in this thesis, Slee and Bownes (1995) confirmed that Drosophila IMPDH is associated with the *raspberry* locus. In addition, they reported that expression of one of the cDNAs in bacteria generates a protein with IMPDH activity. Neither Sifri *et al.* (1994) nor Slee and Bownes (1995) presented a genomic DNA sequence analysis.

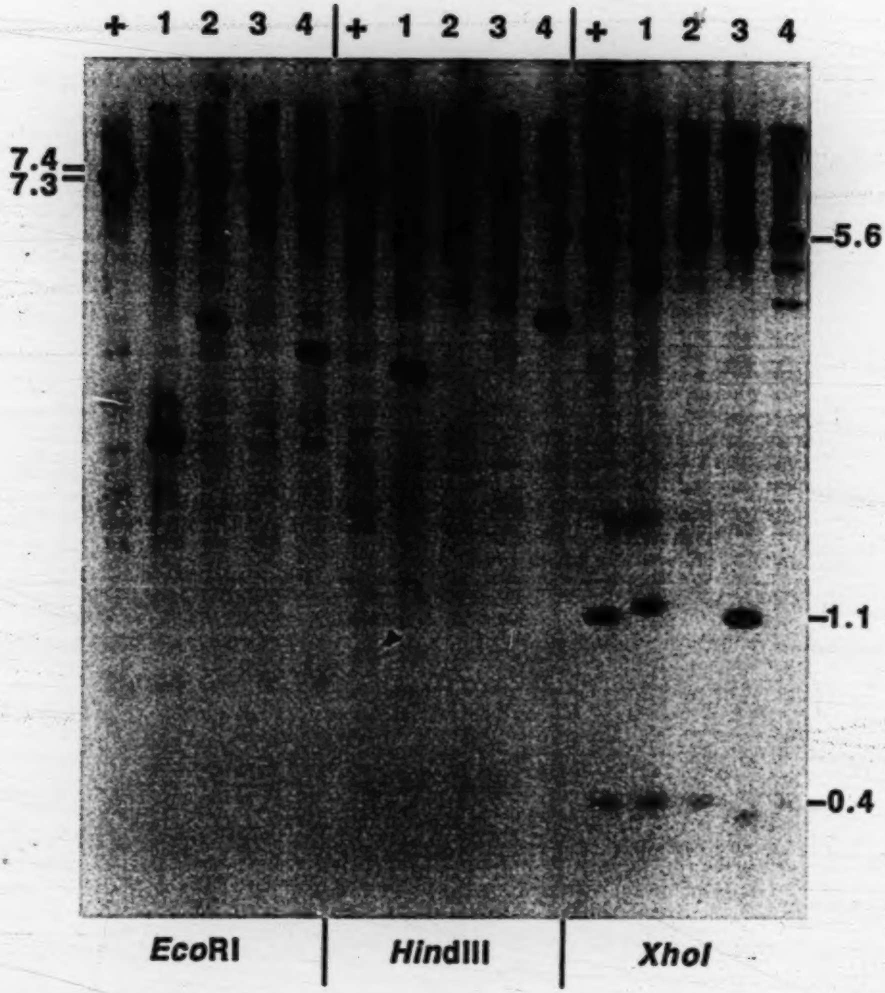
In their Southern analysis of the *ras* eye-colour mutants, Slee and Bownes (1995) obtained mostly the same results as in this thesis, that is to say, all the DNA alterations are in the 4.0kb *Eco* RI-*Hin* dIII region. However, there is a difference in the interpretation of the molecular lesion in *ras*^{3}. Slee and Bownes claimed that *ras*^{3} was an insertional mutant with a complex restriction pattern, instead of a deletion. They suggested that one of the insertions in *ras*^{3} was a 1kb insert inserted in the 5.5kb *Xho* I fragment at the 3′ of the IMPDH gene. Since this claim conflicts with the finding on *ras*^{3} in this thesis, which shows clearly that *ras*^{3} contains a small deletion in the 0.4kb *Xho* I fragment, it is necessary to compare the evidence from Southern analyses presented in the paper of Slee and Bownes (1995) and in this thesis.

For convenience, the data from both Southern analyses and the restriction map of *ras* locus constructed by Slee and Bownes are presented here (Figure 4-1, Figure 4-2, Figure 4-3). The probe used in Slee and Bownes' paper was their 7.5kb *Eco* RI fragment, which is equivalent to the 7.3kb *Eco* RI fragment in this thesis. The probe used in this thesis was the 4.0kb *Eco* RI-*Hin* dIII fragment (4 EH in the diagram), which is part of their larger 7.5kb fragment.

The analysis of the Southern results presented by Slee and Bownes (1995) raises a number of questions. First, the Southern figure appears to be a mixture, possibly derived from several different experiments. The *Xho* I digestion is the only one clearly generated on a single gel. In both *Hin* dIII and *Pst* I digestions, the lanes of *ras*^{4}, possibly as well as the wild-type, were most likely taken from separate gels because they are different in appearance from the others in the same digestion. Second, where they show apparently contiguous lanes in the Xho I digestion, the samples ran unevenly. Third, the size of some fragments, especially the large ones, in their experiment is different from ours and is very difficult to interpret (see below).

One of the two major confusions in the Southern analysis of Slee and Bownes is the *Hin* dIII digestion pattern. In wild-type DNA, their 7.5kb *Eco* RI probe should pick up two *Hin* dIII fragments, one 7.5kb containing the suggested entire *ras* transcribed sequences (the 4.0kb *Eco* RI-*Hin* dIII fragment of the probe plus about 3.4kb upstream sequences), and one 8.0kb containing the presumably non-transcribed downstream sequences (the 3.3kb *Hin* dIII-*Eco* RI fragment of the probe plus about 4.7kb downstream sequence, see Figure 4-3). The *Hin* dIII digested wild-type DNA in Figure 4-2 shows a broad band which can be supposed to represent these two bands. If the only DNA alterations in *ras*^{1}, *ras*^{2} and *ras*^{4} are in the 4.0kb *Eco* RI-*Hin* dIII fragment as mentioned by Slee and Bownes (1995), the size of the 3′ end 8.0kb *Hin* dIII fragment in these mutants should be unchanged and the smaller *Hin* dIII-*Eco* RI portion of this probe (the 3.3kb *Hin* dIII-*Eco* RI fragment) should hybridize to the same 8.0kb *Hin* dIII

Figure 4-1 Southern analysis of Am^{+} and *ras* viable mutants



This is the same figure as Figure 3-12.

Hybridization of ^{32}P-labeled 4.0kb *Eco* RI-*Hin* dIII DNA to Southern-transferred digests of genomic DNA from wild-type (+) and *ras1-4* (1-4) adult females. All flies are homozygous at the *ras* locus. The size marks indicate, on the left, the wild-type *Eco* RI (7.3kb) and *Hin* dIII (7.4kb) fragments and, on the right, the visible wild-type *Xho* I fragments. Long exposure used to visualize the 0.4kb *Xho* I fragment reveals low-level hybridization elsewhere on the autoradiograph. The origin of this background signal is clarified in Figure 3-13. The background obscures potential visualization of an ~11kb *Xho* I fragment (with ~0.2kb of homology with the 4.0kb probe) located upstream of the locus. The fragment is unaltered in the mutants. (Taken from Nash et al., 1994)

Page 130 has been removed due to copyright restrictions. The information removed was Figure 4-2 Southern analysis of Am$^+$ and *ras* viable mutants (Slee and Bownes) taken from Slee and Bownes, 1995.

Page 131 has been removed due to copyright restrictions. The information removed was Figure 4-3 Restriction map of *ras* locus (Slee and Bownes) taken from Slee and Bownes, 1995.

fragment in all three samples. However, the size of the *Hin* dIII bands (about 7.5-8.0kb) in Slee and Bownes' Southern analysis are apparently different in *ras*^{1}, *ras*^{2} and *ras*^{4}.

The situation in Slee and Bownes (1995) is even more surprising regarding *ras*^{3}; both the 7.5kb and 8.0kb *Hin* dIII fragments are missing and are substituted by a much larger double band. In our experiments, the 7.4kb *Hin* dIII band (equivalent to their 7.5kb fragment) is present in the *ras*^{3} samples (Figure 4-1). Our 7.4kb *Hin* dIII fragment may run a little slower than the equivalent wild-type band but it also contains less DNA. Because of this loading difference, the size of the *ras*^{3} *Hin* dIII fragment should be considered as unchanged. We have not studied *ras*^{3} with the 3.3kb *Hin* dIII-*Eco* RI fragment.

The other major confusion occurs with the *Xho* I digestion pattern. Slee and Bownes (1995) reported that *ras*^{3} was an insertional mutant principally based on the analysis of the *Xho* I digestion. As noted above, the samples in the *Xho* I digestion ran unevenly in Slee and Bownes' analysis, with the left part running slower than the right part. The 0.4kb *Xho* I bands in *ras*^{1} and *ras*^{2} ran slower than the 0.4kb band in wild-type, which was in the right-most lane, and the equivalent band in *ras*^{3}, which was in between *ras*^{2} and the wild-type, ran at about the same position as the wild-type. Even though the 0.4kb fragment in *ras*^{1} and *ras*^{2} ran slower than the wild-type, Slee and Bownes claimed that the size of 0.4kb band in both *ras*^{1} and *ras*^{2} is normal because of the unevenly running gel. They also noted that the size of the 0.4kb band in *ras*^{3} is reduced, which agrees with our results.

Considering the unevenly running gel, the sizes of the other two *Xho* I fragments (1.1kb and 5.5kb) in *ras*^{3} are most probably the same as the equivalent wild-type fragments, although they ran slower. It is not clear why Slee and Bownes claimed that the size of the 5.5kb *Xho* I fragment in *ras*^{3} was increased and ignored their observation that the size of the 0.4kb *Xho* I fragment of *ras*^{3} is reduced in the interpretation of their own Southern analysis.

In addition to the *ras*^{3} samples, the size of the 5.5kb *Xho* I fragment in the other three *ras* mutants, *ras*$^{1,\,2\text{ and }4}$, is clearly different in their Southern analysis (Figure 4-2). However, the size of this band is clearly the same as the equivalent wild-type fragment in *ras*$^{1,\,2\text{ and }4}$ in our experiments (Figure 4-1). No explanation was provided by Slee and Bownes in their paper.

Based on above analysis, there are not enough data in Slee and Bownes' paper to support the claim that *ras*^{3} is an insertional mutant. In contrast, the small deletion in the 0.4kb *Xho* I fragment of *ras*^{3} was evident in both sets of data and our sequence analysis confirms the reality of the deletion. Thus, *ras*^{3} is most reasonably interpreted as a deletion mutant.

Slee and Bownes (1995) agree with our analysis of *ras*$^{1,\,2\text{ and }4}$. However, they present no reasonable explanation for all the confusing observations in their *Hin* dIII and *Xho* I digestions and it is difficult to generate one from their data. It is also difficult to understand why they ignored completely several major anomalies found in their digestions.

4.1.2 IMPDH and the *ras* complementation group

This work effectively demonstrates that IMPDH is encoded by the *ras* locus. IMPDH is the first of the two enzymes required for GMP synthesis from IMP, the product of purine *de novo* biosynthesis, and is the rate-limiting enzyme for GMP production in mammals (Weber, 1983). IMPDH converts IMP to XMP, which is then converted to GMP by GMP synthetase (see Figure 1-3).

In the study of IMPDH gene expression in wild-type flies and three *ras* eye-colour mutants, *ras*$^{1,\,2\text{ and }3}$, the preponderance of data indicates the levels of *ras* transcripts and IMPDH activity are low in the mutants during the third and fourth days of pupal life, when eye pigmentation occurs. In *ras*^{3}, the enzyme activity and the level of transcript in males is about half of that in females. However, the enzyme activity in *ras*^{3} females, which show wild-type eye-colour, is the same or even lower than that in *ras*^{1} females, which show mutant eye-colour. No satisfactory explanation is available for this conflicting observation. No direct evidence is available to demonstrate that the gene expression is low in the mutant eye tissues since the data were generated from the whole animals. Nevertheless, the findings suggest that a deficiency of IMPDH activity during the period of eye pigmentation is probably the primary reason for the *ras* eye-colour mutant phenotype. It is instructive to examine the genetics of the entire *ras* complex in terms of alterations of IMPDH activity, even though, as we shall see, the involvement of other gene products will undoubtedly prove to be important.

The genetics of the *ras* locus is complex (Janca *et al.*, 1986). Three groups of mutants are involved. The first is the rare spontaneous red-brown eye-colour *ras* viable mutants (*ras*$^{1-4}$) which are defective in eye-pigment pteridine synthesis (Nolte, 1959). The second group includes three auxotrophs. The first one, *gua1*^{ts}, is a guanosine auxotroph at the restrictive temperature (29°C). The other two, *pur1*^{1} and *pur1*^{2}, are allelic and can be supplemented by either guanosine or adenosine. *pur1*^{2} responds equally well to both guanosine and adenosine supplementation, but *pur1*^{1} responds to guanosine better than adenosine (Falk and Nash, 1974b). The third group contains recessive lethal mutations, the *raspberry-lethals (ras-l)*. These lethal mutants are the mutations most commonly recovered from the *ras* locus, implying that some general function of the *ras* locus is essential (Janca *et al.*, 1986).

The three auxotrophs were originally assigned to two loci, *guanosine1 (gua1)* and *purine1 (pur1)*, because *gua1*^{ts} complements both *pur1*^{1} and *pur1*^{2} (Falk and Nash, 1974b). The difference between *pur1*^{1} and *pur1*^{2} is also reflected in the complementation of *gua1*^{ts}. Although the double heterozygote *gua1*^{ts}/*pur1*^{1} can survive on unsupplemented medium under restrictive conditions, it develops slowly. The developmental delay can be corrected by guanosine supplementation. This suggests that the molecular lesion in *pur1*^{1} probably not only damaged the *pur1* function, but also somehow affected the *gua1* function. Thus, Falk and Nash (1974b) suggested that there are probably two closely linked genes involved. Johnson *et al.* (1979) mapped both *pur1* and *gua1* cytologically to the region 9E1-9E4, the same location as the viable eye-colour mutant *raspberry (ras)*. Complementation studies revealed that *ras*^{2} complements *gua1*^{ts}, *pur1*^{1} and *pur1*^{2} regarding both the auxotrophic phenotype and mutant eye-colour. Based on this observation, despite partial failure of complementation between *gua1*^{ts} and *pur1*^{1}, *ras, gua1* and *pur1* were initially defined as three individual genes (Johnson *et al.*, 1979).

The *ras-l* mutants form a single lethal complementation group. It is the *ras-l* mutants that connect *ras* eye-colour mutants to the *gua1* and *pur1* mutants since many of the lethals fail to complement all three (Janca *et al.*, 1986). Because of this complementation data, *gua1*, *pur1* and *ras* can be thought of as falling into a single functional complex, which has been referred to as the *ras* complex. Alteration of the *ras* complex would, then, be the source of *ras* eye-colour mutants, the auxotrophs and recessive lethal mutants.

The guanosine auxotroph *gua1*^{ts} is explicable by a mutation of the IMPDH gene. An auxotroph that can be rescued by exogenous guanosine is most probably defective in the conversion of IMP to GMP. Since this reaction is catalyzed by IMPDH and by GMP synthetase, reduction of IMPDH activity caused by a *ras* gene mutation could certainly affect GMP production, producing a guanosine auxotroph. Total loss of IMPDH activity and the subsequent deficiency of GMP biosynthesis might also explain the *ras-l* mutants. Because HGPRT activity is absent in *Drosophila melanogaster*, neither hypoxanthine nor guanine can be salvaged (Becker, 1974a; Johnson *et al.*, 1980b, c). Although guanosine can be salvaged, presumably through a kinase activity, the net utilization of guanosine is low, compared with, for example, adenosine (Johnson *et al.*, 1980a, b, c). Thus, were *de novo* GMP biosynthesis blocked completely, it would seem reasonable to suppose that the inefficient salvage synthesis of guanine nucleotides in flies would prove insufficient to permit survival. It is probable, then, that IMPDH activity is an essential function of the *ras* locus, and possibly the only one.

The *ras-l* mutants do not usually (if at all) survive until pupariation (Nash *et al.*, 1981). This contrasts with the observation that lethal mutants in the *de novo* purine biosynthesis pathway (*Prat*$^-$, *ade2*$^-$ and *ade3*$^-$) die mostly in the pupal life and some of them can even survive eclosion (Tiong *et al.*, 1989; Tiong and Nash, 1990; Clark, 1994). GMP biosynthesis is probably affected in both cases: that is to say, due to the lack of IMP from *de novo* purine biosynthesis or due to the loss of IMPDH activity in *ras-l* mutants. However, in the mutants with the blocked *de novo* purine biosynthesis, GMP can still be synthesized from exogenous adenine sources. This would not be the case in *ras-l* mutants with their postulated defect in IMPDH activity, since the conversion uses the IMPDH/GMPS pathway. Thus, loss of the capacity of *de novo* GMP biosynthesis probably has more severe ramifications than the loss of the capacity of *de novo* purine biosynthesis.

It has recently been found that recessive lethal mutants from the *burgundy* locus, which encodes GMP synthetase, normally survive to pupal life (Nash, personal communication). This observation appears to contradict the previous argument that mutants with blocked *de novo* GMP biosynthesis would rarely survive to pupal life. However, the *bur-l* mutants behave peculiarly in several ways, making it difficult to insist that the observations can be taken to refute the argument.

Altered IMPDH activity can also be invoked to explain the phenotype of the red-brown eye-colour *ras* viable mutants. The mutant eye-colour is caused by reduction of the pteridine eye-pigment drosopterin. Pteridines are derived from GTP. IMPDH activity appears to be low in the *ras* mutants during eye pigmentation. Presumably, low IMPDH activity generates low levels of GTP, and subsequently, low levels of pteridines. Mutants with blocked purine *de novo* biosynthesis also show a similar red-brown eye-colour, which is probably the result of similar low GTP levels (Tiong *et al.*, 1989; Tiong and Nash, 1990).

Evans and Howells (1978) reported that the activity of guanosine triphosphate cyclohydrolase (GTPCH) is low in *ras* mutants. The reduction of GTPCH activity is most significant in the heads of late pupae and young adults, when eye-colour formation occurs. GTPCH is the first and rate-limiting enzyme in pteridine biosynthesis from GTP. Low GTPCH activity would be expected to result in low levels of pteridines, leading to mutant eye-colour. Since GTPCH is encoded at the *Punch* locus (Mackay and O'Donnell, 1983), the low GTPCH activity in *ras* mutants must be a secondary effect.

Most pteridine biosynthesis is used for production of pigment or excretion of excess purines (see Nash and Henderson, 1982). With a limited guanine nucleotide pool, pteridine synthesis might well be curtailed to ensure a sufficient supply of purines for important functions. Hatakeyama *et al.* (1992) reported that in IMPDH inhibitor (tiazofurin, 2-amino-1,3,4-thiadiazole and mycophenolic acid) treated mammalian cells, not only was the concentration of cellular GTP reduced, but also was the tetrahydrobiopterin content, which was used as an indicator of GTPCH activity. The tetrahydrobiopterin content can be restored by adding free GTP to the cultures. Therefore, according to the authors, intracellular GTP concentration appears likely to regulate GTPCH activity. However, there are some concerns. First, GTP, along with other phosphorylated nucleotides, cannot be freely transferred through the cell membrane. Therefore, the intracellular GTP may not have a direct influence of GTPCH activity. Second, even if the free GTP can get into the cell for some reason, the proposed restoration of GTPCH function might be simply a matter of the availability of the enzyme substrate, without any regulatory implications.

We have suggested that a parallel response occurs in *ras* mutants and in other mutants in purine biosynthesis *(ade2*$^-$*, ade3*$^-$*, Prat*$^-$*, bur*$^-$*)* which display the same eye-colour phenotype (Nash *et al.*, 1994). Although the mammalian effect appears to be the result of a direct effect on enzyme activity with limited substrate, the balance of the evidence suggests a regulatory effect on enzyme production in Drosophila, since Evans and Howells (1978) assayed levels of GTPCH activity in protein extracts under standard substrate conditions.

The purine auxotrophs in the *ras*-complex cannot be explained as easily as the *gua1*, *ras-1* and *ras* phenotypes. The fact that an auxotroph can be rescued by either purine source conflicts with our understanding of purine metabolism in *Drosophila melanogaster*: Guanine sources are not generally expected to rescue auxotrophs with blocked AMP biosynthesis, nor adenine sources to rescue blocked GMP biosynthesis. Given the lack of GMP reductase activity in *Drosophila melanogaster*, guanine sources cannot, either, be expected to rescue auxotrophs with blocked *de novo* IMP biosynthesis, such as *ade2* and *ade3* mutants (Henikoff *et al.*, 1986c). It is then a puzzle that the *pur1* auxotrophs can be supplemented by either an adenine or a guanine source (Nash and Henderson, 1982).

A possible explanation of the *pur1* phenotype can be postulated as the following: The molecular lesion in *pur1* auxotrophs lowers, but does not completely eliminate, the substrate affinity of the mutated IMPDH. The catalytic function of the enzyme is not affected by the mutation. Since the mutant IMPDH cannot bind IMP properly, little reaction can take place under normal circumstances. An increased concentration of substrate (IMP) might disproportionally elevate IMPDH activity and correct the defect by raising the residual binding efficiency to a higher level. The supposition is that the active site of the mutant IMPDH remains intact, so that once IMPDH binds IMP, the reaction can take place in a normal manner. Since both *pur1* auxotrophs are partial loss-of-function alleles of the *ras* gene (Johnson *et al.*, 1979), if the *pur1* auxotrophs are truly defective in IMPDH, residual activity of IMPDH must be assumed to present, although not enough to allow dietary purine-free survival. When exogenous adenine sources are present, AMP can be synthesized from these exogenous sources in the *pur1* auxotrophs, so that intracellular IMP derived from either *de novo* biosynthesis or conversion of adenine sources would become available for GMP biosynthesis. Of course, when exogenous guanosine is present, the deficiency of IMPDH activity will not be a problem.

One objection to this hypothesis remains: the $gua1^{ts}$ and $pur1^{2}$ alleles appear to complement, which would seem contrary to the hypothesis that all *ras* mutants are defective in IMPDH function. However, because IMPDH functions as a dimer (Shimura *et al.*, 1983; Hupe *et al.*, 1986; Verham *et al.*, 1987), complementation between $pur1^{2}$ and $gua1^{ts}$ could be explained by inter-allelic complementation. Inter-allelic complementation can occur between two distinct types of alleles, neither of which can, by itself, give rise to any appreciable enzyme activity. In the heteroallelic situation, the interaction between the two mutant enzyme subunits (heterodimer) may compensate each other and restore, or partially restore, enzyme function (Crick and Orgel, 1964). According to this hypothesis, the heterodimer of $pur1^{2}/gua1^{ts}$ can provide sufficient normal function if the mutations in $pur1^{2}$ and $gua1^{ts}$ affect different sub-functions of IMPDH, presumably the substrate binding site in $pur1^{2}$ and the active site in $gua1^{ts}$. In the case of $pur1^{1}$, in addition to *pur1* function, the mutation perhaps also partially affects the *gua1* function (the catalytic function). Since all three auxotrophs were induced by EMS, this explanation seems applicable, although certainly not proven.

4.2 Possible models for regulation of the *ras* gene

Independent of molecular data, it is worthwhile examining the available genetic information to limit the possible models for regulation of expression of the *ras* gene (*ras*-complex). As mentioned above, the *ras* gene is a house-keeping gene which encodes an essential function; alterations of its function generate three kinds of mutations, viable eye-colour mutants, auxotrophs and recessive lethal mutants. The *ras* larval eye disks express mutant eye colour autonomously after being transplanted into wild-type hosts (Beadle and Ephrussi, 1936). Tiong (unpublished results) has noticed that no patches of homozygous *ras-l* tissues were ever recovered in the induced mitotic recombination experiments. However, the very high frequency of morphological alterations recovered from these experiments implies that the induced *ras-l* homozygous cells are formed, and survive for several cell generations, then die: according to Tiong, the morphology of the abnormal structures formed corresponds with that found in experiments where tissue patches are known to be killed. The data suggest, then, that homozygous epidermal cells die prematurely, although not immediately.

Casual examination of the phenotypes of the auxotrophs and eye-colour mutants suggests that these mutants do not fall into a single graded series of hypomorphs. An example of complex genetic data that can be explained as a series of hypomorphic mutants is found for the *rudimentary (r)* gene, which encodes a protein catalyzing the first three steps in *de novo* pyrimidine biosynthesis. According to Falk and Nash (1974a), a set of 20 *r* mutants, all of them showing some kind of pyrimidine auxotrophic phenotype, can be regarded as a single series of hypomorphic alleles of the *r* gene. These mutants could be categorized into four groups based on the exogenous pyrimidine requirement, wing shape and female fertility: A) Poor survival even with pyrimidine supplement, rudimentary wings and female sterility; B) Normal survival with pyrimidine supplement, less extreme rudimentary wings and female sterility; C) Pyrimidine auxotrophs, normal wings and normal female fertility; and D) Temperature-sensitive auxotrophs (requiring supplement at 29°C), normal wings and normal female fertility. The severity of the defect in these mutants can be expressed on a single scale with the strongest hypomorphs (and, presumably, amorphs) in the group A and the weakest hypomorphs in the group D. If the situation of the *ras* complex was as simple as that, the defects in the auxotrophs should be severer than, or at least the same as, that in eye-colour mutants. However, the surviving auxotrophs show wild-type eye-colour on diets that yield mutant eye-colour in the classic *ras* mutants. Combined with the fact that all the eye-colour mutants are fully viable, even on restrictive purine diets, both types of mutants cannot be regarded as falling into the same graded hypomorphic series.

The behavior of one *ras-l* mutant, *ras*^{l19} (J5), which yields occasional escapers, also conflicts with the graded hypomorph hypothesis. The eye-colour of the rare survivors of *ras*^{l19} is clearly closer to wild-type than to that of the viable *ras* mutants. If the graded activity hypothesis were valid, the eye-colour of the homozygous and hemizygous survivors carrying the lethal allele *ras*^{l19} would be expected to be at least as extreme as that of the *ras* viable mutants.

Janca *et al.* (1986) rejected the graded hypomorphic hypothesis based on complementation results. In a systematic complementation study of *ras*^{2}, *gua1*^{ts}, *pur1*^{1}, *pur1*^{2} against a group of eight EMS induced *ras-l* mutants, Janca *et al.* (1986) demonstrated that the differences between the *ras-l* mutants are allele-specific as defined by their phenotypes found in double heteroallelic combinations (Table 4-1). In a second test, Nash (unpublished data) has shown that the same *ras-l* allele-specific differences are also present in the double heterozygous combination against *ras*^{1}, *ras*^{3} and *ras*^{4}. The most telling evidence presented by Janca *et al.* (1986) is that two fully lethal mutants, *ras*^{l15} (F1) and *ras*^{l16} (F19) complement the *ras* eye-colour mutants by producing wild-type eye-colour in the relevant heteroallelic combinations. Because of this, *ras*^{l15} and *ras*^{l16} have previously been referred to as *(ras)*-lethal (*(ras)*^{l}) mutants. If the lethal mutants really exhibited greater deficiency of IMPDH activity in all tissues than did the *ras* viable mutants, the heterozygotes could not produce higher IMPDH activity in the eyes than that produced in the viable *ras* eye-colour mutants.

Interactions between heteroallelic dimers might explain the allele-specific phenomena in complementation results (*Section 4.1.2*). However, this kind of complementation mostly results from amino acid substitution in protein subunits (Crick and Orgel, 1964). It does not fit the mode of complementation against *ras* eye-colour mutants since the molecular lesions in all of them are located in non-coding regions. In addition, because there is no question of complementation in the *ras*^{l19} escapers, the closer to normal eye-colour of the escapers cannot be explained by interaction between heterodimers.

Table 4-1 A summary of the expression of *ras-l* mutations in heterozygous combinations with *gua1*^{ts}, *pur1*^{1}, *pur1*^{2} and *ras*^{2}

	Non-lethal mutation			
ras-l allele	*gua1*^{ts}	*pur1*^{1}	*pur1*^{2}	*ras*^{2}
ras^{l14} (D12)	Strong	Strong	Weak	Strong
ras^{l15} (F1)	Strong	Strong	Weak	Weak
ras^{l16} (F19)	Strong	Strong	Weak	Weak
ras^{l17} (H6)	Strong	Strong	Strong	Strong
ras^{l18} (H25)	Strong	Strong	Weak	Strong
ras^{l19} (J5)	Weak	Strong	Strong	Strong
ras^{l20} (K27)	Weak	Strong	Strong	Strong
ras^{l21} (N23)	Strong	Strong	Weak	Strong

To simplify the situation, the phenotype of the heterozygote is categorized either as "strong" or "weak". (Modified from Janca *et al.*, 1986)

There are several possible explanations for how wild-type eye-colour was produced in double heterozygotes of *ras*^{115}/*ras* viable and *ras*^{116}/*ras* viable mutants: first, wild-type eye-colour is the result of interaction of the products of the two alleles (for example, the heterodimer situation described before); second, the eye-colour is generated from the activity of the viable allele as a result of an interaction with the lethal allele or its products; and third, the eye-colour results from activity of the lethal allele in cells where the viable allele is insufficiently active.

As discussed previously, inter-allelic complementation between heterodimers is not likely the case since the mutations in the viable mutants do not affect the coding sequences. The second possibility implies that the viable allele may function normally by interacting directly with either the lethal allele or its products. This interaction would correct the malfunction of the viable allele (rather than its products). However, since the molecular lesions in *ras*$^{1,\ 2\ \mathrm{and}\ 4}$ are large intronic insertions and in *ras*^{3} is a deletion upstream of the transcriptional start site, it seems unlikely that two different lethal alleles containing presumed point mutations can correct both insertional and deletional mutations.

The third possibility seems more attractive. It suggests that in double heterozygotes, the wild-type eye-colour is provided by the lethal allele. This implies that even though the lethal alleles cannot generate enough IMPDH activity to allow survival, it can provide enough IMPDH activity for eye-colour formation. In contrast, even though the viable allele provides sufficient function to support survival, it cannot generate that required for eye-pigment formation. This indicates that the eye-colour function of the *ras* gene is regulated independently from the viability function. Hence, Nash *et al.* (1994) suggested that the *ras* gene probably is ubiquitously expressed (house-keeping function), but tissue-specifically regulated (tissue-specific function). The house-keeping function is expressed in all tissues and required for viability. The tissue-specific function is individually regulated in specific tissues at specific developmental stages depending on specific requirements.

The occurrence of two kinds of diametrically opposed mutant phenotypes -- the viable *ras* eye-colour viable mutants and the *(ras)-l* mutants *ras*$^{115\ \mathrm{and}\ 116}$, in which the eye-colour function appears unaffected, strongly supports the hypothesis. These two kinds of mutants imply that the tissue-specific and house-keeping functions of the *ras* gene can be mutated separately; some components in the *ras* gene regulatory apparatus appear responsible only for the eye-colour function, and others only for house-keeping functions. The hypothesis, then, can be further expanded to encompass the idea that both the *ras* viable and the *(ras)*-lethal mutants are regulatory mutants.

In fact, although this concept is attractive, it is not logically necessary. Either the *ras* or the *(ras)-l* mutants could be examples of enzyme defects falling onto a continuum from normal to completely null enzyme activity, but not both. Since the molecular lesions in the *ras* viable eye-colour mutants are located either in the 5′ untranscribed region or in an intron and do not affect the coding region, the *ras* mutants are most likely regulatory mutants. The *(ras)-l* mutants could also be regulatory mutants, as discussed previously. They probably have lost, or, reduced house-keeping function but retain sufficient tissue-specific eye-colour function so they can complement *ras* viable mutants. However, there is no definite evidence to support this idea. Although the *(ras)-l* mutants complement *ras* viable mutants with respect to the eye-colour, they could still be defective in enzyme activity if eye-colour formation requires lower enzyme activity than viability.

Based on the house-keeping/tissue-specific regulation hypothesis, most of the *ras* lethal mutants must be deficient in both aspects of *ras* function. For example, in *ras*^{l24}, IMPDH transcription probably is abolished because of the alternative transcriptional initiation and termination introduced by the P-element. Since there is no *ras* function in the *ras*-lethal mutants, they cannot complement *ras* viable mutants in the heteroallelic situation. The mutation in *ras*^{l19}, whose rare survivors show closer to normal eye-colour than found in *ras*$^{1 \text{ and } 2}$ (Nash *et al.*, 1994), probably seriously impedes the house-keeping function but is much less effective in reducing the eye specific function. The mutation could be defective either in house-keeping regulatory functions or enzyme activity. The auxotrophs could be regarded as enzyme defective mutations, as discussed previously.

The *ras* eye-colour spontaneous mutants are rare, only four having been recorded in almost a century. In the search for *ras* mutant alleles, three EMS induced auxotrophic alleles, eight EMS induced recessive lethal alleles, about nine transposon element induced recessive lethal alleles and a number of X-ray induced recessive alleles have been recovered. However, no induced eye-colour viable mutant was found. This observation implies that the supposed eye-specific regulatory elements of the *ras* gene are extremely limited, and the supposed house-keeping specific regulatory elements seem to be more extensive and/or mutable, since two out of eight lethal mutants probably lost only their house-keeping functions, if they are truly regulatory mutants.

4.3 *cis*-acting regulatory elements in *ras* gene

4.3.1 *cis*-acting intronic splicing regulatory sequence

Molecular analysis has revealed that three of the four *ras* viable mutants, *ras*^{1}, *ras*^{2} and *ras*^{4}, are insertional mutations. All three inserts are located in a 12bp region of the second intron of the *ras* gene. This finding strongly supports the genetic analysis that the components of the tissue-specific regulatory mechanism are extremely limited. The fourth one, *ras*^{3}, in which, quite unusually, the mutant eye-colour is present only in the males, contains a 37bp deletion well upstream of the transcriptional start site. Thus, it appears likely that two sequences involved in regulation of the eye-colour function of the *ras* gene have been identified in this study. The fact that *ras*^{3} is allelic to the insertional mutants suggests that the two regulatory sequences can be formally considered as *cis*-acting regulatory sequences.

Sequencing of part of the three inserts has revealed that they are different. All three inserts contain long-terminal-repeats and are accompanied by short direct genomic repeats at the insertion sites, indicating their retrotransposon origin (Corces and Geyer, 1991; Gierl and Frey, 1991). Transposon insertions often cause mutations in *Drosophila melanogaster* either by introducing new regulatory signals, resulting in alternative transcriptional initiation and/or premature termination or by destroying the integrity and function of host DNA elements, such as regulatory and/or coding sequences. Northern data show no obvious reduction of IMPDH transcripts in the young adults of *ras*$^{1, 2 \text{ and } 4}$. However, there is an additional large poly-A^{-} IMPDH transcript present in these mutants, as well as some very weak intermediate transcripts. The size of the large transcripts is about the same as the IMPDH mRNA plus the insert in each case. This suggests that the entire *ras* gene, including the inserts, can be transcribed in these mutants. Because of the potential full length of the transcripts (IMPDH transcript plus insert), similar mutant eye-colour, almost identical insertion site and the different content of the inserts, it is most likely that interruption of the integrity and function of the 12bp insertion site in the second intron, not the property of the inserts, is the primary cause of the rare *ras* mutant phenotype.

Genetic analysis suggests that *ras* viable mutants are defective in tissue-specific regulation of IMPDH. Based on genetic data, the 12bp intronic sequence most likely is a tissue-specific regulatory sequence. It could be either a tissue-specific transcriptional regulatory sequence whose function is required for initiation of the *ras* gene transcription in the eyes, or it could be a tissue-specific regulatory sequence which is required for post-transcriptional regulation in the eyes. Northern results showed that the *ras* gene is transcribed efficiently in young adults of *ras*$^{1,\ 2 \text{ and } 4}$. Since the data were obtained from the whole body, they cannot be taken to refute the possibility that eye specific transcriptional initiation of the *ras* gene is abolished or reduced by the insertion. However, because of the presence of the large, unspliced transcripts, the simplest explanation is that the 12bp sequence is involved in pre-mRNA processing and forms part of an intronic splicing enhancer. This presumed intronic splicing enhancer might be the binding site of the splicing proteins (snRNPs and/or SR proteins) for precisely recognizing the intron/exon boundaries during RNA splicing (Watakabe et al., 1993; Reed, 1996). Thus, insertions in this site might prevent, or partially prevent, the proper binding of the pre-mRNA and the splicing proteins. As a result, normal RNA processing might not be performed and the intron, along with the insert, is retained in some pre-mRNAs, resulting in unspliced RNA transcripts. Of course, it is also possible that the 12bp sequence is not required for splicing, if the sequence contains any regulatory functions at all, and the unspliced transcripts are just the result of the physical change of the integrity of the intron.

Experimentally it has not been established whether the splicing deficiency presents in all tissues or only in the eyes during eye-pigmentation, since the current data were obtained from whole animals. If unspliced transcripts accumulated in all tissues, insertion in this 12bp sequence and the subsequently reduced splicing efficiency would seem to have no major effect on the vital function of the *ras* gene. However, the more interesting possibility is that it occurs specifically during eye-colour formation in the eyes. Northern analysis revealed that, in *ras*^{1} and *ras*^{2}, there are proportionally more unspliced transcripts in the third day of pupal life than at any other stage. In addition, enzyme activity analysis showed that in these mutants, the IMPDH activity is particularly low in the late pupal stage. This leads to the possibility that the low IMPDH activity could be the result of the large amount of eye-specific unspliced IMPDH transcript, due to low splicing efficiency, in the eyes. This hypothesis can be expanded to suggest that there may be eye-specific splicing mechanisms involving the proposed intronic splicing enhancer.

The observation of high IMPDH activity in young adults of *ras*^{4} is a surprise. Since a deficiency of IMPDH activity probably is the primary cause of these eye-colour mutants, IMPDH activity was not expected to be increased in them. One explanation could be that if the insertion does result in a reduction of IMPDH activity in the eyes, the lack of enzyme activity in the eyes might turn out to be a signal of IMPDH deficiency to the flies. As the result, the flies may produce more IMPDH in other parts of the body to compensate for the lack of IMPDH in the eye. This may lead to an over-production of IMPDH in the whole body, resulting in a high level of IMPDH activity. However, this cannot explain the lower level of IMPDH activity in *ras*^{1} and *ras*^{2} which have a similar molecular lesion.

4.3.2 *cis*-acting dosage compensating regulatory element

The remaining *ras* viable mutant, *ras*^{3}, behaves differently from the other three *ras* mutants. In *ras*^{3}, the mutant eye-colour is expressed in hemizygous males but not in homozygous females. Genetic analysis revealed that the eye-colour in hemizygous *ras*^{3}/deficiency females is very similar to the mutant eye-colour in hemizygous *ras*^{3} males. In *Drosophila melanogaster*, most genes located on the single X chromosome in males are "up regulated" or "dosage compensated" to produce levels of products equivalent to that synthesized from the two X chromosomes in females. The simplest explanation of the *ras*^{3} male specific phenotype is that *ras*^{3} males lack the capacity for dosage compensation. Preliminary analysis of enzyme activity shows that, in late pupal life, IMPDH activity in *ras*^{3} males is about half of the level of *ras*^{3} females, which agrees with the genetic analysis.

Females hemizygous for the wild-type *ras* gene (*ras*$^{+}$*/Df ras*$^{-}$) are wild-type with normal eye-colour. Since *ras*^{3}*/Df ras*$^{-}$ flies exhibit a mutant eye-colour, it follows that *ras*^{3} must also be hypomorphic. There is no evidence suggesting that the defects in *ras* function caused by the *ras*^{3} mutant extend to house-keeping functions, but this cannot be ruled out.

Dosage compensation is achieved by doubling the transcriptional initiation rate of X-linked genes in males (Jaffe and Laird, 1986). Numerous genetic studies have indicated that dosage compensation in the male Drosophila, like a great deal of transcriptional regulation, is achieved by the interactions between the *trans*-acting transcriptional regulatory proteins and the multiple *cis*-acting specific DNA sequences along the length of X-chromosome (Jaffe and Laird, 1986; Lucchesi and Manning, 1987). The dosage compensation regulatory proteins form a transcriptional complex only in males (Bashaw and Baker, 1996) because, in females, the production of components of the complex is inhibited by the female specific Sxl protein (Bashaw and Baker, 1997; Kelley *et al.*, 1997). The dosage compensation complex binds to multiple dosage compensation DNA sequences along the X-chromosome and up-regulates the transcription efficiency of X-linked genes (Kuroda *et al.*, 1991; Palmer *et al.*, 1993; Baker *et al.*, 1994). This mechanism ensures that dosage compensation occurs only in the males.

The molecular lesion in *ras*^{3} is a 37bp deletion in a region with exceptionally high GC content located upstream of the transcriptional initiation site of the IMPDH gene. As mentioned above, the mutation in *ras*^{3} abolishes dosage compensation of the IMPDH gene in males, so that this 37bp sequence is extremely likely to be involved in dosage compensation at the level of assembly of the transcription regulatory complex on the male X chromosome.

The P-element in *ras*^{124} is inserted right at the 3′ end of the deletion in *ras*^{3}. It seems that dosage compensation may also be abolished in *ras*^{124} because the integrity of the dosage compensation element probably was also destroyed by the insertion. However, the genetics of *ras*^{124} is totally different from *ras*^{3}. *ras*^{124} is a recessive lethal mutation. Northern analysis revealed that an ~1.0kb transcript from the

P-element is present in heterozygous *ras*^{D4}*/FM6* flies (Figure 3-5). These data suggest that the insertion of the P-element interrupts *ras* transcription by introducing alternative transcriptional initiation and termination signals.

4.3.3 Imperfect trinucleotide (CAG) repeat sequences

An imperfect CAG trinucleotide repeat sequence is found in the 3′ untranslated region, 57bp downstream from the putative translational stop codon in cDNA29. The sequence contains 48bp with a total of twelve CAGs, including a run of nine consecutive CAG repeats. The sequence is represented in our genomic sequence. In cDNA83, there is one less (i.e. eight) CAG repeat in the run. IMPDH cDNA sequences reported by both Sifri *et al.* (1994) and Slee and Bownes (1995) contain the same imperfect CAG repeat sequence except with only seven consecutive CAG repeats. Thus, the number of CAG repeats varies in different cDNAs. The origin of this variation is not yet defined. It might result from genomic variation, transcriptional slippage, or even be a cloning effect. Tandem repeats with variable numbers of glutamine codons (CAG or CAA) are present in several genes in Drosophila (Michalakis and Veuille, 1996). If the imperfect trinucleotide sequence in cDNA29 were translated in appropriate frame, the CAG repeats would yield an amino acid sequence consisting mainly of glutamines interrupted by two histidines. A similar amino acid sequence is found in some proteins and is known as an OPA motif (Wharton et al., 1985).

Our Southern analysis revealed that there are multiple sequences which hybridize to this imperfect trinucleotide sequence in the Drosophila genome (Figure 3-13). The function of these repeat sequences is unknown. Indeed, there is a large (~200bp) sequence containing many such trinucleotide repeats located ~1.5kb upstream from the putative transcriptional start site, although it is certainly not organized strictly as an OPA-like region. The two imperfect trinucleotide sequences may be potential binding sites for regulatory proteins, because of their high GC contents. It is even possible that they have the capacity to form a loop structure in the establishment of chromatin organization, thereby affecting transcription. We have no further evidence regarding this conjecture, however.

In the past few years, trinucleotide tandem repeats have been identified as a source of several human hereditary diseases, such as spinal and bulbar muscular atrophy (SBMA), Huntington's disease, dentatorubral-pallidoluysian atrophy (DRPLA), Machado-Joseph disease (MJD), fragile X syndrome, myotonic dystrophy and Friedreich's ataxia (Tsuji, 1997). The size (number of trinucleotides) of trinucleotide repeats is relatively low in the wild-type genome. It is the expansion of the trinucleotide runs which causes the diseases. The size of expanded repeats is highly unstable. It is often different from parents to offspring and even is different between different tissues in the same patient (La Spada, 1997). The location of these repeats can be in coding regions as well as non-coding regions (Reddy and Housman, 1997). Molecular study has revealed that the size of the repeats is positively related to the severity of the disease (Reddy and Housman, 1997). The fact that the expansion of trinucleotide repeats located in non-coding regions (5′ or 3′ untranslated region or intron) can cause diseases suggests that these trinucleotide repeats may have some regulatory functions.

4.4 Oogenesis and the *cis*-acting intronic splicing enhancer

In the three insertional mutants, the females of *ras*^{1} and *ras*^{2} are fertile but *ras*^{4} females are sterile. The escapers of *ras*^{119}/*ras*^{119} are also sterile and so are the double heterozygous female *ras*^{4}/*ras*^{119}, suggesting that the infertility is caused by the lack of an intrinsic function of the *ras* gene.

The fertility of *ras*^{1} and *ras*^{2} seems to contrast strongly with the infertility of *ras*^{4}. It seems likely, however, that the contrast is relatively trivial. The ovaries of *ras*$^{1,\ 2\ \text{and}\ 4}$ were dissected. The oocytes in *ras*^{4} are underdeveloped. The size of the largest oocytes in the ovarioles is only about half of the normal length. In *ras*^{1} and *ras*^{2}, many of the ovarioles are apparently normal, but some appear similar to those in *ras*^{4} (Nash, personal communication). In addition, in these two mutants, the most unexpected discovery is that IMPDH RNA is absent from the early embryos. Since it is virtually certain that the RNA normally found in early embryos is a maternal product, it appears that even in *ras*$^{1\ \text{and}\ 2}$ there is a very significant effect on oogenesis. It seems that the maternal IMPDH transcripts are dispensable, for both *ras*^{1} and *ras*^{2} embryos are viable and appear to develop at a normal rate. This brings into question the role of the message in embryogenesis.

There are several possible explanations of the lack of maternal transcripts (either spliced or unspliced) in *ras*^{1} and *ras*^{2} embryos. One is that no IMPDH transcripts are synthesized during oogenesis. Northern data showed that there is no lack of IMPDH transcripts in young adult females of *ras*^{1}, *ras*^{2} and *ras*^{4}. Ironically, *ras*^{4} females probably produce more IMPDH RNA than wild-type female flies. However, since no studies have yet been directed exclusively at the ovary, the possibility that no IMPDH RNA is synthesized in the ovaries of these three mutants cannot be ruled out. Given that there is evidence of a failure of proper pre-mRNA processing, another possibility is that the IMPDH transcripts are formed (presumably in the nurse cells) but do not reach the oocytes, perhaps because normal mRNA is required for active mRNA transportation from nurse cells to oocytes. Of course, it is also possible that the absence of transcripts is due to the mutant RNA being degraded either before or after transfer to oocytes. There is little doubt that, no matter which possibility turns out to be correct, further study of this curious phenomenon could be most rewarding.

4.5 Conclusion

The function of the *ras* gene is complex. Genetic studies reveal that *ras* function is required for viability (probably basic house-keeping functions), as well as for eye-colour formation during the period of eye-pigmentation and oocyte development (tissue-specific function). House-keeping and tissue-specific functions probably are regulated, at least to some degree, by independent regulatory elements. The tissue-specific regulatory elements are extremely limited. In contrast, the house-keeping specific regulatory elements are apparently more extensive. The *ras* eye-colour mutants probably are defective in tissue-specific regulation and the *(ras)*^{l} mutants are defective in specific house-keeping regulation.

Sequence analysis reveals that IMPDH, the rate limiting enzyme in the GMP biosynthesis pathway, is encoded at the *ras* locus. Thus the principle function of the *ras* gene is probably mediating GMP biosynthesis. The phenotypes of all mutants in the *ras* complex, *ras* eye-colour mutants, auxotrophs and lethal mutants, can be rationalized as results of altered IMPDH activity.

The *ras* gene is a typical house-keeping gene. There is no TATA-box in the 5′ upstream regulatory region. Instead there is an initiator-like element and an extremely high GC content sequence in the upstream region. The combination of initiator and GC rich sequences often is related to house-keeping regulation.

Two other *cis*-acting regulatory sequences seem to have been identified. One, probably a splicing enhancer, is located in the second intron. Insertions in this intronic element generate the classical *ras* eye-colour mutants, and, somehow, affect oocyte development. The other *cis*-acting sequence is a dosage compensation regulatory sequence located upstream of the transcriptional initiation site. In addition, there is a CAG trinucleotide repeat (OPA sequence) located in the 3′ untranslated region and another region containing CAG repeats located in the 5′ upstream region, which could conceivably have regulatory significance.

4.6 Future plans

This work has revealed some important molecular aspects of the *ras* gene. However, several questions remain to be answered, three of which seem most important. First, the molecular lesions of other classes of *ras* mutants need to be characterized, in order to understand the regulation of the gene and the functioning of its product. Secondly, the *ras* eye-colour mutants need to be further studied and the two *cis*-acting elements suggested by these mutants need to be defined, in order to understand their roles in the gene expression, particularly in the eyes. Last, the expression of the *ras* gene in the ovaries needs to be studied, in order to analyze the production and deposition of the maternal transcript and to address the absence of maternal transcript in the early embryos of *ras*^{1} and *ras*^{2}.

The *ras* gene is complex. Investigation of its mutants could provide important information on molecular aspects of the gene. Except for the four spontaneous *ras* eye-colour mutants and one P-element induced lethal mutant, the molecular lesions in the *ras* mutants have not yet been defined. The auxotrophs and most of the lethal mutants were induced by EMS, so that they are probably point mutations. To pinpoint such mutations requires sequencing; currently the best approach is to utilize PCR amplification and SSCP (single-strand conformation polymorphism) to identify a limited region to be sequenced, starting from a panel of PCR primers which spans the entire gene.

Characterization of the mutants may clarify some aspects of *ras* function, and, quite probably, reveal new regulatory elements and mechanisms. Accordingly, new questions and new experiments might be appropriate. The present tissue-specific regulatory model may be confirmed or a new one suggested by the results. The key issue regarding to the current model is the nature of the two lethal mutants, which complement *ras* eye-colour mutants. Clearly, if upon molecular definition they appear likely to be regulatory mutants, the tissue-specific regulatory model would be supported. If they are not, the whole hypothesis may need to be re-evaluated. A generally useful tool in the investigation of the functioning of lethals would involve checking the maternal transcripts produced by females heteroallelic for *ras*^{1} (or *ras*^{2})

and the lethals. The absence or presence of transcripts (and, in the latter case, the nature of the transcripts) would help define the effect of a lethal and, in the case of the two mutants, perhaps confirm them as exhibiting a regulatory defect.

To further clarify the nature of the eye-colour mutants, the developmental expression profile of the *ras* gene in these mutants needs to be further defined. Study of the RNA transcript and the enzyme activity in dissected parts of the flies collected from different developmental stages could provide valuable information. To increase the sensitivity, RT-PCR could be used in RNA experiments. *In situ* histochemical investigation, with either nucleotide or antibody as probe, can reveal (and even quantify), the pattern of the gene expression. Intron and the insert sequences could be used as probes to identify the distribution and relative levels of the large transcripts in $ras^{1,\ 2\ and\ 4}$.

The upstream regulatory sequence identified in ras^3 is probably involved in dosage compensation. Dosage compensation has been extensively studied. However, no *cis*-acting DNA element that regulates dosage compensation of a single gene has yet been defined. Study of the upstream sequence might lead to the identification of a dosage compensation regulatory element. The upstream sequence (with or without the 37bp region deleted in ras^3) could be inserted in front of a reporter gene in a transposon and introduced into host flies. If sex differential expression of the reporter gene occurs with the wild-type transposon but expression is not sexually differentiated with the deletion, a test system for examination of the dosage compensation would be established. The dosage compensation sequence could then be further defined using *in vitro* mutagenesis and transgenic analysis. In addition, the upstream sequence could be tested for binding to the dosage compensation complex or its sub-components.

The 12bp intronic insertion target site sequence identified in $ras^{1,\ 2\ and\ 4}$ may be involved in RNA splicing, probably in a tissue-specific matter. Study of this element could provide insights on tissue-specific regulation. The transgenic approach using a reporter gene might be used to test the regulatory function of the sequence, but the approach would be more complex: The cause of the unspliced transcripts is unknown. Since the unspliced transcript only counts for a small portion of the total IMPDH transcripts, the splicing efficiency seems to be reduced by the insertion. The reduced efficiency could be a general effect of the enlarged intron size because of the insertions, or it could be the result of the interruptions of the integrity of the 12bp insertion site, which might be required for splicing in specific tissues. In addition, multiple transcripts might be generated and/or special splicing mechanism might be used in order to fulfill different functions of the gene, especially in different tissues. With all the unresolved concerns, it is very difficult to design a suitable test system incorporating a reporter gene.

A general approach of transgenically rescuing *ras* mutants with wild-type gene could be used to study molecular aspects of the *ras* gene. This methodology would permit a transgenic approach to understanding the 12bp intronic sequence. The second intron of ras^2 could be used to substitute the wild-type one to see if the ras^2 mutant phenotype can be reproduced in the host flies. If the ras^2 phenotype can be reproduce, the system is ready for use in *in vitro* mutagenesis and transgenic experiments. Convenient reporter systems might then be devised to improve the experimental efficiency.

There is essentially no known explanation for the absence of maternal IMPDH transcript in the early embryos of *ras*^{1} and *ras*^{2}. It is not even clear if identical IMPDH transcript is produced in both eyes and ovaries. In general, wild-type transcripts that are found in mature oocytes are thought to be produced in the nurse cells and transferred to the oocyte; of cause, there is no specific information on the *ras* gene. As discussed earlier, it is possible that the transcript either is not produced or cannot be deposited in the mutant oocytes. *In situ* hybridization in ovaries would address this question. If the gene is not transcribed in ovaries, further experiments might be designed to find out if the intronic sequence is involved in transcriptional initiation in ovaries. If the gene is transcribed, the transfer mechanism, which the mutant phenotype suggests to be receptor mediated, should be analyzed. If the research exposes the involvement of a system of active transportation of maternal transcripts, it may lead to identification of a significant new aspect of oogenesis.

Bibliography

Adams, M. D., D. Z. Rudner, and D. C. Rio. 1996. Biochemistry and regulation of pre-mRNA splicing. *Current Opinion in Cell Biology* 8 (3): 331-9.

Agarwal, V. R., C. I. Ashanullah, E. R. Simpson, and S. E. Bulun. 1997. Alternatively spliced transcripts of the aromatase cytochrome P450 (CYP19) gene in adipose tissue of women. *Journal of Clinical Endocrinology & Metabolism* 82 (1): 70-4.

Allison, L. A., M. Moyle, M. Shales, and C. J. Ingles. 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* 42 (2): 599-610.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215 (3): 403-10.

Arndt, K. T., C. Styles, and G. R. Fink. 1987. Multiple global regulators control HIS4 transcription in yeast. *Science* 237 (4817): 874-80.

Baker, B. S. 1973. The maternal and zygotic control of development by cinnamon, a new mutant in Drosophila melanogaster. *Developmental Biology* 33 (2): 429-40.

Baker, B. S., M. Gorman, and I. Marin. 1994. Dosage compensation in Drosophila. *Annual Review of Genetics* 28: 491-521.

Barbash, D. A., and T. W. Cline. 1995. Genetic and molecular analysis of the autosomal component of the primary sex determination signal of Drosophila melanogaster. *Genetics* 141 (4): 1451-71.

Bashaw, G. J., and B. S. Baker. 1996. Dosage compensation and chromatin structure in Drosophila. *Current Opinion in Genetics & Development* 6 (4):496-501.

Bashaw, G. J., and B. S. Baker. 1997. The regulation of the Drosophila msl-2 gene reveals a function for Sex-lethal in translational control. *Cell* 89 (5): 789-98.

Bashkin, P., and O. Sperling. 1978. Some regulatory properties of purine biosynthesis de novo in long-term cultures of epithelial-like rat liver cells. *Biochimica et Biophysica Acta* 538 (3): 505-11.

Beadle, G. W., and B. Ephrussi. 1936. The differentiation of eye pigments in Drosophila as studied by transplantation. *Genetics* 21: 225-247.

Becker, J. L. 1974a. Purine metabolism pathways in Drosophila cells grown "in vitro" : Phosphoribosyl transferase activities. *Biochimie* 56 (5): 779-81.

Becker, J. L. 1974b. Purine metabolism in Drosophila melanogaster cells in culture in vitro: purine interconversion. *Biochimie* 56 (9): 1249-53.

Becker, J. L. 1975. Purine metabolism in cultured Drosophila melanogaster cells: adenosine utilization. *Comptes Rendus Hebdomadaires des Seances de l Academie des Sciences - D: Sciences Naturelles* 281 (16): 1171-4.

Bell, L. R., E. M. Maine, P. Schedl, and T. W. Cline. 1988. Sex-lethal, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* 55 (6): 1037-46.

Belote, J. M. 1989. The control of sexual development in Drosophila melanogaster: genetic and molecular analysis of a genetic regulatory hierarchy--a minireview. *Gene* 82 (1): 161-7.

Berg, O. G., and P. H. von Hippel. 1988. Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. *Journal of Molecular Biology* 200 (4): 709-23.

Berget, S. M. 1995. Exon recognition in vertebrate splicing. *Journal of Biological Chemistry* 270 (6): 2411-4.

Berget, S. M., C. Moore, and P. A. Sharp. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* 74 (8): 3171-5.

Bird, A. P. 1995. Gene number, noise reduction and biological complexity. *Trends in Genetics* 11 (3): 94-100.

Blake, C. C. F. 1978. Do genes-in-pieces imply proteins-in-pieces? *Nature* 273: 267.

Blake, C. 1983. Exons--present from the beginning? *Nature* 306: 535-537.

Boggs, R. T., P. Gregor, S. Idriss, J. M. Belote, and M. McKeown. 1987. Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene. *Cell* 50 (5): 739-47.

Bohme, E., and G. Schultz. 1974. Separation of cyclic nucleotides by thin-layer chromatography on polyethyleneimine cellulose. *Methods in Enzymology* 38: 27-38.

Bokal, A. J., W. Ross, T. Gaal, R. C. Johnson, and R. L. Gourse. 1997. Molecular anatomy of a transcription activation patch: FIS-RNA polymerase interactions at the Escherichia coli rrnB P1 promoter. *EMBO Journal* 16 (1): 154-62.

Bonifer, C., M. C. Huber, U. Jagle, N. Faust, and A. E. Sippel. 1996. Prerequisites for tissue specific and position independent expression of a gene locus in transgenic mice. *Journal of Molecular Medicine* 74 (11): 663-71.

Bopp, D., G. Calhoun, J. I. Horabin, M. Samuels, and P. Schedl. 1996. Sex-specific control of Sex-lethal is a conserved mechanism for sex determination in the genus Drosophila. *Development* 122 (3): 971-82.

Brack, C., and S. Tonegawa. 1977. Variable and constant parts of the immunoglobulin light chain gene of a mouse myeloma cell are 1250 nontranslated bases apart. *Proceedings of the National Academy of Sciences of the United States of America* 74 (12): 5652-6.

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72: 248-54.

Breathnach, R., and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annual Review of Biochemistry* 50: 349-83.

Breathnach, R., J. L. Mandel, and P. Chambon. 1977. Ovalbumin gene is split in chicken DNA. *Nature* 270 (5635): 314-9.

Brown, G. M. 1971. The biosynthesis of pteridines. *Advances in Enzymology & Related Areas of Molecular Biology* 35: 35-77.

Brunekreef, G. A., H. J. Kraft, J. G. Schoenmakers, and N. H. Lubsen. 1996. The mechanism of recruitment of the lactate dehydrogenase-B/epsilon-crystallin gene by the duck lens. *Journal of Molecular Biology* 262 (5): 629-39.

Bulun, S. E., and E. R. Simpson. 1994. Regulation of aromatase expression in human tissues. *Breast Cancer Research & Treatment* 30 (1): 19-29.

Buratowski, S., S. Hahn, L. Guarente, and P. A. Sharp. 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 56 (4): 549-61.

Burg, A. W., and G. M. Brown. 1966. The biosynthesis of folic acid. VI. Enzymatic conversion of carbon atom 8 of guanosine triphosphate to formic acid. *Biochimica et Biophysica Acta* 117 (1): 275-8.

Burg, A. W., and G. M. Brown. 1968. The biosynthesis of folic acid. 8. Purification and properties of the enzyme that catalyzes the production of formate from carbon atom 8 of guanosine triphosphate. *Journal of Biological Chemistry* 243 (9): 2349-58.

Burley, S. K., and R. G. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annual Review of Biochemistry* 65: 769-99.

Busby, S., and R. H. Ebright. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79 (5): 743-6.

Bussey, H., D. B. Kaback, W. Zhong, D. T. Vo, M. W. Clark, N. Fortin, J. Hall, B. F. Ouellette, T. Keng, A. B. Barton et al. 1995. The nucleotide sequence of chromosome I from Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* 92 (9): 3809-13.

Carlo, T., D. A. Sterner, and S. M. Berget. 1996. An intron splicing enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon. *RNA* 2 (4): 342-53.

Cavalier-Smith, T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *Journal of Cell Science* 34: 247-78.

Cavalier-Smith, T. 1991. Intron phylogeny: a new hypothesis. *Trend in Genetics* 7 (5): 145-148.

Cavener, D. R. 1987. Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. *Nucleic Acids Research* 15 (4): 1353-61.

Chabot, B. 1996. Directing alternative splicing: cast and scenarios. *Trends in Genetics* 12 (11): 472-8.

Chen, S., P. L. Nagy, and H. Zalkin. 1997. Role of NRF-1 in bidirectional transcription of the human GPAT-AIRC purine biosynthesis locus. *Nucleic Acids Research* 25 (9): 1809-16.

Chomey, E., and D. Nash. 1995. The guanosine monophosphate synthetase gene of Drosophila melanogaster. 39th Annual Meeting of the Genetics Society of Canada, Guelph, Ontario.

Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* 162 (1): 156-9.

Chow, L. T., R. E. Gelinas, T. R. Broker, and R. J. Roberts. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12 (1): 1-8.

Christiansen, L. C., S. Schou, P. Nygaard, and H. H. Saxild. 1997. Xanthine metabolism in Bacillus subtilis: characterization of the xpt-pbuX operon and evidence for purine- and nitrogen-controlled expression of genes involved in xanthine salvage and catabolism. *Journal of Bacteriology* 179 (8): 2540-50.

Clark, D. V. 1994. Molecular and genetic analyses of Drosophila *Prat*, which encodes the first enzyme of *de nove* purine biosynthesis. *Genetics* 136 (2): 547-557.

Clark, D. V., and S. Henikoff. 1992. Unusual organizational features of the Drosophila Gart locus are not conserved within Diptera. *Journal of Molecular Evolution* 35 (1): 51-9.

Cline, T. W. 1984. Autoregulation functioning of a Drosophila gene product that establish es and maintains the sexually determined state. *Genetics* 107 (2): 231-77.

Cline, T. W. 1993. The Drosophila sex determination signal: how do flies count to two? *Trends in Genetics* 9 (11): 385-90.

Collado-Vides, J., B. Magasanik, and J. D. Gralla. 1991. Control site location and transcriptional regulation in Escherichia coli. *Microbiological Reviews* 55 (3): 371-94.

Collart, F. R., and E. Huberman. 1988. Cloning and sequence analysis of the human and Chinese hamster inosine-5'-monophosphate dehydrogenase cDNAs. *Journal of Biological Chemistry* 263 (30): 15769-72.

Corces, V. G., and P. K. Geyer. 1991. Interactions of retrotransposons with the host genome: the case of the gypsy element of Drosophila. *Trends in Genetics* 7 (3): 86-90.

Counce, S. J. 1957. A female sterile mutant (deep orange) of Drosophila melanogaster increasing isoxanthopterin content. *Experientia* 13: 354.

Crick, F. H. C., and L. E. Orgel. 1964. The theory of inter-allelic complementation. *J. Mol. Biol.* 8: 161-165.

Czarniecki, D., R. J. Noel Jr, and W. S. Reznikoff. January 1997. The -45 region of the Escherichia coli lac promoter: CAP-dependent and CAP-independent transcription. *Journal of Bacteriology* 179 (2): 423-9.

Daignan-Fornier, B., and G. R. Fink. 1992. Coregulation of purine and histidine biosynthesis by the transcriptional activators BAS1 and BAS2. *Proceedings of the National Academy of Sciences of the United States of America* 89 (15): 6746-50.

de Crombrugghe, B., S. Busby, and H. Buc. 1984. Cyclic AMP receptor protein: role in transcription activation. *Science* 224 (4651): 831-8.

Dibb, N. J. 1993. Why do genes have introns? *Federation of European Biochemical Societies* 325 (1,2): 135-139.

Dickson, R. C., J. Abelson, W. M. Barnes, and W. S. Reznikoff. 1975. Genetic regulation: the Lac control region. *Science* 187 (4171): 27-35.

Doolittle, W. F. 1978. Genes in pieces: were they ever together? *Nature* 272: 581-582.

Dorsett, D., and K. B. Jacobson. 1982. Purification and biosynthesis of quench spot, a drosopterin precursor in Drosophila melanogaster. *Biochemistry* 21 (6): 1238-43.

Douglas, S. E. 1995. DNA Strider. An inexpensive sequence analysis package for the Macintosh. *Molecular Biotechnology* 3 (1): 37-45.

Dreyfuss, G., M. J. Matunis, S. Pinol-Roma, and C. G. Burd. 1993. hnRNP proteins and the biogenesis of mRNA. *Annual Review of Biochemistry* 62: 289-321.

Dufau, M. L., C. H. Tsai-Morris, Z. Z. Hu, and E. Buczko. 1995. Structure and regulation of the luteinizing hormone receptor gene. *Journal of Steroid Biochemistry & Molecular Biology* 53 (1-6): 283-91.

Duffy, J. B., and J. P. Gergen. 1991. The Drosophila segmentation gene runt acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene Sex-lethal. *Genes & Development* 5 (12A): 2176-87.

Dunn, L. C. 1934. Mutant characters. *Drosophila Information Service* 1: 30.

Dynan, W. S. 1986. Promoters for housekeeping genes. *Trends in Genetics* 2 (8): 196-197.

Dynan, W. S. 1989. Modularity in promoters and enhancers. *Cell* 58 (1): 1-4.

Ebbole, D. J., and H. Zalkin. 1987. Cloning and characterization of a 12-gene cluster from Bacillus subtilis encoding nine enzymes for de novo purine nucleotide synthesis. *Journal of Biological Chemistry* 262 (17): 8274-87.

Ebbole, D. J., and H. Zalkin. 1988. Detection of pur operon-attenuated mRNA and accumulated degradation intermediates in Bacillus subtilis. *Journal of Biological Chemistry* 263 (22): 10894-902.

Ebbole, D. J., and H. Zalkin. 1989. Bacillus subtilis pur operon expression and regulation. *Journal of Bacteriology* 171 (4): 2136-41.

Eick, D., A. Wedel, and H. Heumann. 1994. From initiation to elongation: comparison of transcription by prokaryotic and eukaryotic RNA polymerases. *Trends in Genetics* 10 (8): 292-6.

Elwood, P. C., K. Nachmanoff, Y. Saikawa, S. T. Page, P. Pacheco, S. Roberts, and K. N. Chung. 1997. The divergent 5' termini of the alpha human folate receptor (hFR) mRNAs originate from two tissue-specific promoters and alternative splicing: characterization of the alpha hFR gene structure. *Biochemistry* 36 (6): 1467-78.

Estes, P. A., L. N. Keyes, and P. Schedl. 1995. Multiple response elements in the Sex-lethal early promoter ensure its female-specific expression pattern. *Molecular & Cellular Biology* 15 (2): 904-17.

Evans, B. A., and A. J. Howells. 1978. Control of drosopterin synthesis in Drosophila melanogaster: mutants showing an altered pattern of GTP cyclohydrolase activity during development. *Biochemical Genetics* 16 (1-2): 13-26.

Falk, D. R. 1973. *Sex-linked auxotrophic mutations in Drosophila melanogaster. Ph.D. Thesis.* Edmonton, Alberta: Department of Genetics, University of Alberta.

Falk, D. R., and D. Nash. 1974a. Pyrimidine auxotrophy in Drosophila melanogaster: normal-winged, auxotrophic mutants and dominant auxotrophy at the rudimentary locus. *Mol. Gen. Genet.* 131 (4): 339-349.

Falk, D. R., and D. Nash. 1974b. Sex-linked auxotrophic and putative auxotrophic mutants of Drosophila melanogaster. *Genetics* 76 (4): 755-66.

Fan, C. L., and G. M. Brown. 1976. Partial purification and properties of guanosine triphosphate cyclohydrolase from Drosophila melanogaster. *Biochemical Genetics* 14 (3-4): 259-70.

Fan, C. L., and G. M. Brown. 1979. Partial purification and some properties of biopterin synthase and dihydropterin oxidase from Drosophila melanogaster. *Biochemical Genetics* 17 (3-4): 351-69.

Fan, C. L., G. G. Krivi, and G. M. Brown. 1975. The conversion of dihydroneopterin triphosphate to sepiapterin by an enzyme system from Drosophila melanogaster. *Biochemical & Biophysical Research Communications* 67 (3): 1047-54.

Fan, C. L., L. M. Hall, A. J. Skrinska, and G. M. Brown. 1976. Correlation of guanosine triphosphate cyclohydrolase activity and the synthesis of pterins in Drosophila melanogaster. *Biochemical Genetics* 14 (3-4): 271-80.

Fischer, G., C. Schmidt, J. Opitz, Z. Cully, K. Kuhn, and E. Poschl. 1993. Identification of a novel sequence element in the common promoter region of human collagen type IV genes, involved in the regulation of divergent transcription. *Biochemical Journal* 292 (Pt 3): 687-95.

Flaks, J., and L. Lukens. 1963. The Enzymes of Purine Nucleotide Synthesis de Novo. *Methods in Enzymology* 6: 52-95.

Flashner, Y., and J. D. Gralla. 1988. Dual mechanism of repression at a distance in the lac operon. *Proceedings of the National Academy of Sciences of the United States of America* 85 (23): 8968-72.

Forrest, H., E. Glassman, and H. K. Mitchell. 1956. Conversion of 2-amino 4-hydroxy pteridine to isoxanthopterin in Drosophila melanogaster. *Science* 124: 725-726.

Fried, M., and D. M. Crothers. 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Research* 9 (23): 6505-25.

Friedman, A. M., T. O. Fischmann, and T. A. Steitz. 1995. Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science* 268 (5218): 1721-7.

Frischauf, A. M., N. Murray, and H. Lehrach. 1987. Lambda phage vectors--EMBL series. *Methods in Enzymology* 153: 103-15.

Frommer, G., R. Schuh, and H. Jackle. April 1994. Localized expression of a novel micropia-like element in the blastoderm of Drosophila melanogaster is dependent on the anterior morphogen bicoid. *Chromosoma* 103 (2): 82-9.

Gallego, M. E., P. Sirand-Pugnet, P. Durosay, B. Clouet d'Orval, Y. d'Aubenton-Carafa, E. Brody, A. Expert-Bezancon, and J. Marie. 1996. Tissue-specific splicing of two mutually exclusive exons of the chicken beta-tropomyosin pre-mRNA: positive and negative regulations. *Biochimie* 78 (6): 457-65.

Gallego, M. E., R. Gattoni, J. Stevenin, J. Marie, and A. Expert-Bezancon. 1997. The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the beta-tropomyosin alternative exon 6A. *EMBO Journal* 16 (7): 1772-84.

Gedvilaite, A., and K. Sasnauskas. 1994. Control of the expression of the ADE2 gene of the yeast Saccharomyces cerevisiae. *Current Genetics* 25 (6): 475-9.

Geer, B. W. 1964. Inheritance of the dietary ribonucleic acid requirement of Drosophila melanogaster. *Genetics* 49: 787-796.

Geiger, J. H., S. Hahn, S. Lee, and P. B. Sigler. 1996. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* 272 (5263): 830-6.

Gierl, A., and M. Frey. 1991. Eukaryotic transposible elements with short terminal inverted repeats. *Curr. Opin. Genet. Dev.* 1 (4): 494-497.

Gilbert, W. 1978. Why genes in pieces? *Nature* 271: 501.

Gilbert, W. 1979. Introns and exons: playgrounds of evolution. *Mol. Cell. Biol.* 14 (ICN-UCLA Symp.): 1-12.

Gilbert, W., and A. Maxam. 1973. The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences of the United States of America* 70 (12): 3581-4.

Gilbert, W., N. Maizels, and A. Maxam. 1974. Sequences of controlling regions of the lactose operon. *Cold Spring Harbor Symposia on Quantitative Biology* 38: 845-55.

Gilbert, W., J. Gralla, J. Majors, and A. Maxam. 1975. Lactose operator sequences and the action of *lac* repressor. Within *Protein-ligand interactions*. H. Sund, and G. Blauer, pp. 93-210. Berlin: de Gruyter.

Go, M. 1981. Correlation of DNA exonic regions with protein structural units in haemoglobin. *Nature* 291 (5810): 90-2.

Goodrich, J. A., and W. R. McClure. 1991. Competing promoters in prokaryotic transcription. *Trends in Biochemical Sciences* 16 (11): 394-7.

Goodrich, J. A., G. Cutler, and R. Tjian. 1996. Contacts in context: promoter specificity and macromolecular interactions in transcription. *Cell* 84 (6): 825-30.

Goto, K., J. L. Heymont, J. Klein-Nulend, H. M. Kronenberg, and M. B. Demay. 1996. Identification of an osteoblastic silencer element in the first intron of the rat osteocalcin gene. *Biochemistry* 35 (33): 11005-11.

Gotta, M., and S. M. Gasser. 1996. Nuclear organization and transcriptional silencing in yeast. *Experientia* 52 (12): 1136-47.

Gralla, J. D. 1989. Bacterial gene regulation from distant DNA sites. *Cell* 57 (2): 193-5.

Gralla, J. D. 1991. Transcriptional control--lessons from an E. coli promoter data base. *Cell* 66 (3): 415-8.

Granadino, B., L. O. F. Penalva, M. R. Green, J. Valcarcel, and L. Sanchez. 1997. Distinct mechanisms of splicing regulation in vivo by the Drosophila protein Sex-lethal. *Proceedings of the National Academy of Sciences of the United States of America* 94 (14): 7343-8.

Green, M. R. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annual Review of Cell Biology* 7: 559-599.

Gregg, T. G., and L. A. Smuckler. 1965. Pteridines and gene homologies in the eye colour of Drosophila hydei and Drosophila melanogaster. *Genetics* 52: 1023-1034.

Grivell, L. A. 1996. Transposition: mobile introns get into line. *Current Biology* 6 (1): 48-51.

Gross, C. A., M. Lonetto, and R. Losick. 1992. Bacterial sigma factors. Within *Transcriptional regulation*. S. L. McKnight, and K. R. Yamamoto, pp. 129-178. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Gu, J. J., J. Spychala, and B. S. Mitchell. 1997. Regulation of the human inosine monophosphate dehydrogenase type I gene. Utilization of alternative promoters. *Journal of Biological Chemistry* 272 (7): 4458-66.

Gumina, R. J., N. E. Kirschbaum, K. Piotrowski, and P. J. Newman. 1997. Characterization of the human platelet/endothelial cell adhesion molecule-1 promoter: identification of a GATA-2 binding element required for optimal transcriptional activity. *Blood* 89 (4): 1260-9.

Hadorn, E., and H. K. Mitchell. 1951. Properties of mutants of Drosophila melanogaster and changes during development as revealed by paper chromatography. *Proc. Natl. Acad. Sci. U.S.A.* 37: 650-655.

Hansen, S. K., and R. Tjian. 1995. TAFs and TFIIA mediate differential utilization of the tandem Adh promoters. *Cell* 82 (4): 565-75.

Hatakeyama, K., T. Harada, and H. Kagamiyama. 1992. IMP dehydrogenase inhibitors reduce intracellular tetrahydrobiopterin levels through reduction of intracellular GTP levels. Indications of the regulation of GTP cyclohydrolase I activity by restriction of GTP availability in the cells. *Journal of Biological Chemistry* 267 (29): 20734-9.

Hawley, D. K., and W. R. McClure. 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Research* 11 (8): 2237-55.

Henderson, J. F. 1972. *Regulation of purine biosynthesis.* ACS Monograph 170: American Chemical Society, Washington, D.C.

Henderson, J. F., and F. W. Scott. 1980. Inhibition of animal and invertebrate cell growth by naturally occurring purine bases and ribonucleosides. *Pharmacology & Therapeutics* 8 (3): 539-71.

Henderson, J. F., J. K. Lowe, and J. Barankiewicz. 1977. Purine and pyrimidine metabolism: pathways, pitfalls and perturbations. *Ciba Foundation Symposium* 48: 3-21.

Henderson, J. F., G. Zombor, and P. W. Burridge. June 1978. Guanosine triphosphate catabolism in human and rabbit erythrocytes: role of reductive deamination of guanylate to inosinate. *Canadian Journal of Biochemistry* 56 (6): 474-9.

Henderson, J. F., F. W. Scott, and J. K. Lowe. 1980. Toxicity of naturally occurring purine deoxyribonucleosides. *Pharmacology & Therapeutics* 8 (3): 573-604.

Hendriks, W., J. W. Mulders, M. A. Bibby, C. Slingsby, H. Bloemendal, and W. W. de Jong. 1988. Duck lens epsilon-crystallin and lactate dehydrogenase B4 are identical: a single-copy gene product with two distinct functions. *Proceedings of the National Academy of Sciences of the United States of America* 85 (19): 7114-8.

Henikoff, S. 1986. The Saccharomyces cerevisiae ADE5,7 protein is homologous to overlapping Drosophila melanogaster Gart polypeptides. *Journal of Molecular Biology* 190 (4): 519-28.

Henikoff, S. 1987. Multifunctional polypeptides for purine de novo synthesis. *Bioessays* 6 (1): 8-13.

Henikoff, S., and M. K. Eghtedarzadeh. 1987. Conserved arrangement of nested genes at the Drosophila Gart locus. *Genetics* 117 (4): 711-25.

Henikoff, S., J. S. Sloan, and J. D. Kelly. 1983. A Drosophila metabolic gene transcript is alternatively processed. *Cell* 34 (2): 405-14.

Henikoff, S., M. A. Keene, K. Fechtel, and J. W. Fristrom. 1986a. Gene within a gene: nested Drosophila genes encode unrelated proteins on opposite DNA strands. *Cell* 44 (1): 33-42.

Henikoff, S., M. A. Keene, J. S. Sloan, J. Bleskan, R. Hards, and D. Patterson. 1986b. Multiple purine pathway enzyme activities are encoded at a single genetic locus in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* 83 (3): 720-4.

Henikoff, S., D. Nash, R. Hards, J. Bleskan, J. F. Woolford, F. Naguib, and D. Patterson. 1986c. Two Drosophila melanogaster mutations block successive steps of de novo purine synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 83 (11): 3919-23.

Henner, D. J., L. Band, and H. Shimotsu. 1985. Nucleotide sequence of the Bacillus subtilis tryptophan operon. *Gene* 34 (2-3): 169-77.

Henry, R. W., C. L. Sadowski, R. Kobayashi, and N. Hernandez. 1995. A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerase II and III. *Nature* 374 (6523): 653-6.

Hernandez, N. 1993. TBP, a universal eukaryotic transcription factor? *Genes & Development* 7 (7B): 1291-308.

Hertel, K. J., K. W. Lynch, E. C. Hsiao, E. H. Liu, and T. Maniatis. 1996. Structural and functional conservation of the Drosophila doublesex splicing enhancer repeat elements. *RNA* 2 (10): 969-81.

Hertel, K. J., K. W. Lynch, and T. Maniatis. 1997. Common themes in the function of transcription and splicing enhancers. *Current Opinion in Cell Biology* 9 (3): 350-7.

Higgins, D. G., and P. M. Sharp. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73 (1): 237-44.

Higgins, D. G., and P. M. Sharp. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *Computer Applications in the Biosciences* 5 (2): 151-3.

Hinton, T., J. Ellis, and D. T. Noyes. 1951. An adenine requirement in a strain of Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* 37: 293-299.

Hodge, L. D., and E. Glassman. 1967. Purine catabolism in Drosophila melanogaster. I. Reactions leading to xanthine dehydrogenase. *Biochim. Biophys. Acta.* 149: 335-343.

Holland, M. J., N. C. Klein, and R. P. Cox. 1978. Experimental modulation of PRPP availability for ribonucleotide synthesis from hypoxanthine in human skin fibroblast cultures. *Experimental Cell Research* 111 (2): 237-43.

Horabin, J. I., and P. Schedl. 1993. Sex-lethal autoregulation requires multiple cis-acting elements upstream and downstream of the male exon and appears to depend largely on controlling the use of the male exon 5' splice site. *Molecular & Cellular Biology* 13 (12): 7734-46.

Horowitz, D. S., and A. R. Krainer. 1994. Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. *Trends in Genetics* 10 (3): 100-6.

Horton, N., M. Lewis, and P. Lu. 1997. Escherichia coli lac repressor-lac operator interaction and the influence of allosteric effectors. *Journal of Molecular Biology* 265 (1): 1-7.

Hoshijima, K., A. Kohyama, I. Watakabe, K. Inoue, H. Sakamoto, and Y. Shimura. 1995. Transcriptional regulation of the Sex-lethal gene by helix-loop-helix proteins. *Nucleic Acids Research* 23 (17): 3441-8.

House, H. L. 1962. Insect nutrition. *Annu. Rev. Biochem.* 31: 653-672.

Hudson, J. M., and M. G. Fried. 1990. Co-operative interactions between the catabolite gene activator protein and the lac repressor at the lactose promoter. *Journal of Molecular Biology* 214 (2): 381-96.

Hultmark, D., R. Klemenz, and W. J. Gehring. 1986. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. *Cell* 44 (3): 429-38.

Hupe, D. J., B. A. Azzolina, and N. D. Behrens. 1986. IMP dehydrogenase from the intracellular parasitic protozoan Eimeria tenella and its inhibition by mycophenolic acid. *Journal of Biological Chemistry* 261 (18): 8363-9.

Iannello, R. C., J. Young, S. Sumarsono, M. J. Tymms, H. H. Dahl, J. Gould, M. Hedger, and I. Kola. 1997. Regulation of Pdha-2 expression is mediated by proximal promoter sequences and CpG methylation. *Molecular & Cellular Biology* 17 (2): 612-9.

Ish-Horowicz, D., S. M. Pinchin, P. Schedl, S. Artavanis-Tsakonas, and M. E. Mirault. 1979. Genetic and molecular analysis of the 87A7 and 87C1 heat-inducible loci of D. melanogaster. *Cell* 18 (4): 1351-8.

Ito, K., and Y. Nakamura. 1996. Localization of nusA-suppressing amino acid substitutions in the conserved regions of the beta' subunit of Escherichia coli RNA polymerase. *Molecular & General Genetics* 251 (6): 699-706.

Jacob, F., and J. Monod. 1961. Genetic regulatory mechanismes in the synthesis of proteins. *J. Mol. Biol.* 3: 318-356.

Jaffe, E., and C. Laird. 1986. Dosage compensation in Drosophila. *Trends in Genetics* 2 (12): 316-21.

Janca, F. C., E. P. Woloshyn, and D. Nash. 1986. Heterogeneity of lethals in a "simple" lethal complementation group. *Genetics* 112 (1): 43-64.

Javahery, R., A. Khachi, K. Lo, B. Zenzie-Gregory, and S. T. Smale. 1994. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Molecular & Cellular Biology* 14 (1): 116-27.

Jensen, E. O., K. Paludan, J. J. Hyldig-Nielsen, P. Jorgensen, and K. A. Marcker. 1981. The structure of a chromosomal leghaemoglobin gene from soybean. *Nature* 291: 677-679.

Jiralerspong, S., and P. I. Patel. 1996. Regulation of the hypoxanthine phosphoribosyltransferase gene: in vitro and in vivo approaches. *Proceedings of the Society for Experimental Biology & Medicine* 212 (2): 116-27.

Johnson, M. M. 1978. *Cytogenetic and biochemical studies on purine requiring mutants of Drosophila melanogaster. M.S. Thesis*. Department of Genetics, University of Alberta: Edmonton, Alberta.

Johnson, M. M., D. Nash, and J. F. Henderson. 1977. Biochemical studies of purine auxotrophs of Drosophila melanogaster. *Advances in Experimental Medicine & Biology* 76A: 110-114.

Johnson, M. M., E. P. Woloshyn, and D. Nash. 1979. Cytogenetic localisation of the purine 1 and guanosine 1 loci of Drosophila melanogaster; the purine 1 locus specifies a vital function. *Molecular & General Genetics* 174 (3): 287-92.

Johnson, M. M., D. Nash, and J. F. Henderson. 1980a. Metabolism of adenosine in larvae of Drosophila melanogaster. *Comp. Biochem. Physiol. B Comp. Biochem.* 66: 549-553.

Johnson, M. M., D. Nash, and J. F. Henderson. 1980b. Purine metabolism in larvae of Drosophila melanogaster fed radioactive hypoxanthine, inosine or formate. *Comp. Biochem. Physiol. B Comp. Biochem.* 66: 555-561.

Johnson, M. M., D. Nash, and J. F. Henderson. 1980c. Metabolism of guanine and guanosine in larvae of Drosophila melanogaster. *Comp. Biochem. Physiol. B Comp. Biochem.* 66: 563-567.

Johnston, M., S. Andrews, R. Brinkman, J. Cooper, H. Ding, J. Dover, Z. Du, A. Favello, L. Fulton, S. Gattung et al. 1994. Complete nucleotide sequence of Saccharomyces cerevisiae chromosome VIII. *Science* 265 (5181): 2077-82.

Johnstone, M. E., and D. Nash. 1979. IMP dehydrogenase levels in a guanosine requiring strain of Drosophila melanogaster. *Genetics* 91: s56.

Johnstone, M. E., D. Nash, and F. N. Naguib. 1985. Three purine auxotrophic loci on the second chromosome of Drosophila melanogaster. *Biochemical Genetics* 23 (7-8): 539-55.

Jones, C. J., and S. Aizawa. 1991. Genetic control of the bacterial flagellar regulon. *Current Opinion in Genetics & Development* 1 (3): 319-23.

Kaiser, K., and M. Meisterernst. 1996. The human general co-factors. *Trends in Biochemical Sciences* 21 (9): 342-5.

Karlik, C. C., and E. A. Fyrberg. May 1985. An insertion within a variably spliced Drosophila tropomyosin gene blocks accumulation of only one encoded isoform. *Cell* 41 (1): 57-66.

Karlin, S., and S. F. Altschul. 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proceedings of the National Academy of Sciences of the United States of America* 87 (6): 2264-8.

Kaufmann, J., and S. T. Smale. 1994. Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes & Development* 8 (7): 821-9.

Keener, J., and M. Nomura. 1993. Dominant lethal phenotype of a mutation in the -35 recognition region of Escherichia coli sigma 70. *Proceedings of the National Academy of Sciences of the United States of America* 90 (5): 1751-5.

Kelley, R. L., J. Wang, L. Bell, and M. I. Kuroda. 8 May 1997. Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism. *Nature* 387 (6629): 195-9.

Kennell, J. C., J. V. Moran, P. S. Perlman, R. A. Butow, and A. M. Lambowitz. 1993. Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell* 73 (1): 133-46.

Keyes, L. N., T. W. Cline, and P. Schedl. 1992. The primary sex determination signal of Drosophila acts at the level of transcription [published erratum appears in Cell 1992 May 1;69(3):following 572]. *Cell* 68 (5): 933-43.

Kim, J., A. Sebring, J. J. Esch, M. E. Kraus, K. Vorwerk, J. Magee, and S. B. Carroll. 1996. Integration of positional signals and regulation of wing formation and identity by Drosophila vestigial gene. *Nature* 382 (6587): 133-8.

Klein, C. B., and M. Costa. 1997. DNA methylation, heterochromatin and epigenetic carcinogens. *Mutation Research* 386 (2): 163-80.

Kohler, J., S. Schafer-Preuss, and D. Buttgereit. 1996. Related enhancers in the intron of the beta1 tubulin gene of Drosophila melanogaster are essential for maternal and CNS-specific expression during embryogenesis. *Nucleic Acids Research* 24 (13): 2543-50.

Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco, and J. L. Manley. 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 368 (6467): 119-24.

Konno, Y., Y. Natsumeda, M. Nagai, Y. Yamaji, S. Ohno, K. Suzuki, and G. Weber. 1991. Expression of human IMP dehydrogenase types I and II in Escherichia coli and distribution in human normal lymphocytes and leukemic cell lines. *Journal of Biological Chemistry* 266 (1): 506-9.

Kornberg, R. D. 1996. RNA polymerase II transcription control. *Trends in Biochemical Sciences* 21 (9): 325-6.

Kraft, H. J., W. Hendriks, W. W. de Jong, N. H. Lubsen, and J. G. Schoenmakers. 1993. Duck lactate dehydrogenase B/epsilon-crystallin gene. Lens recruitment of a GC-promoter. *Journal of Molecular Biology* 229 (4): 849-59.

Krivi, G. G., and G. M. Brown. April 1979. Purification and properties of the enzymes from Drosophila melanogaster that catalyze the synthesis of sepiapterin from dihydroneopterin triphosphate. *Biochemical Genetics* 17 (3-4): 371-90.

Kuroda, M. I., M. J. Kernan, R. Kreber, B. Ganetzky, and B. S. Baker. 1991. The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila. *Cell* 66 (5): 935-47.

Kustu, S., A. K. North, and D. S. Weiss. 1991. Prokaryotic transcriptional enhancers and enhancer-binding proteins. *TIBS* 16: 397-402.

La Spada, A. R. 1997. Trinucleotide repeat instability: genetic features and molecular mechanisms. *Brain Pathology* 7 (3): 943-63.

Lalli, E., J. S. Lee, M. Lamas, K. Tamai, E. Zazopoulos, F. Nantel, L. Penna, N. S. Foulkes, and P. Sassone-Corsi. 1996. The nuclear response to cAMP: role of transcription factor CREM. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 351 (1336): 201-9.

Lamas, M., L. Monaco, E. Zazopoulos, E. Lalli, K. Tamai, L. Penna, C. Mazzucchelli, F. Nantel, N. S. Foulkes, and P. Sassone-Corsi. 1996. CREM: a master-switch in the transcriptional response to cAMP. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 351 (1339): 561-7.

Lee, C. G., P. D. Zamore, M. R. Green, and J. Hurwitz. 1993. RNA annealing activity is intrinsically associated with U2AF. *Journal of Biological Chemistry* 268 (18): 13472-8.

Lennard, A. C., and M. Fried. 1991. The bidirectional promoter of the divergently transcribed mouse Surf-1 and Surf-2 genes. *Molecular & Cellular Biology* 11 (3): 1281-94.

Leonard, N. J. 1986. *Cloning the raspberry locus of Drosophila melanogaster: a locus involved in purine metabolism. M.S. Thesis,*. Edmonton, Alberta: Department of Genetics, University of Alberta.

Levine, M., and T. Hoey. 1988. Homeobox proteins as sequence-specific transcription factors. *Cell* 55 (4): 537-40.

Lewis, M., G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271 (5253): 1247-54.

Li, M., W. R. McClure, and M. M. Susskind. 1997. Changing the mechanism of transcriptional activation by phage lambda repressor. *Proceedings of the National Academy of Sciences of the United States of America* 94 (8): 3691-6.

Lucchesi, J. C. 1996. Dosage compensation in Drosophila and the "complex' world of transcriptional regulation. *Bioessays* 18 (7): 541-7.

Lucchesi, J. C., and T. Skripsky. 1981. The link between dosage compensation and sex differentiation in Drosophila melanogaster. *Chromosoma* 82 (2): 217-27.

Lucchesi, J. C., and J. E. Manning. 1987. Gene dosage compensation in Drosophila melanogaster. *Advances in Genetics* 24:371-429.

Lynch, K. W., and T. Maniatis. 1996. Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. *Genes & Development* 10 (16): 2089-101.

Ma, D., I. Olave, A. Merino, and D. Reinberg. 1996. Separation of the transcriptional coactivator and antirepression functions of transcription factor IIA. *Proceedings of the National Academy of Sciences of the United States of America* 93 (13): 6583-8.

Mackay, W. J., and J. M. O'Donnell. 1983. A genetic analysis of the pteridine biosynthetic enzyme, guanosine triphosphate cyclohydrolase, in Drosophila melanogaster. *Genetics* 105 (1): 35-53.

Madhani, H. D., and C. Guthrie. 1994. Dynamic RNA-RNA interactions in the spliceosome. *Annual Review of Genetics* 28: 1-26.

Maeda, T., B. H. Devens, S. Fukuse, C. Turck, and D. R. Webb. 1995. Evidence for a calcium regulated, bidirectional intronic promoter in the murine TCR V alpha 1 gene. *International Immunology* 7 (8): 1339-51.

Magoulas, C., and D. A. Hickey. 1992. Isolation of three novel Drosophila melanogaster genes containing repetitive sequences rich in GCN triplets. *Genome* 35 (1): 133-9.

Mandel, J. L. 1993. Questions of expansion. *Nature Genetics* 4 (1): 8-9.

Maniatis, T., S. Goodbourn, and J. A. Fischer. 1987. Regulation of inducible and tissue-specific gene expression. *Science* 236 (4806): 1237-45.

Marck, C. 1988. 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Research* 16 (5): 1829-36.

Matthews, K. S. 1996. The whole lactose repressor. *Science* 271 (5253): 1245-6.

McClure, W. R. 1985. Mechanism and control of transcription initiation in prokaryotes. *Annual Review of Biochemistry* 54: 171-204.

McKay, D. B., I. T. Weber, and T. A. Steitz. 1982. Structure of catabolite gene activator protein at 2.9-A resolution. Incorporation of amino acid sequence and interactions with cyclic AMP. *Journal of Biological Chemistry* 257 (16): 9518-24.

McKnight, S. L. 1982. Functional relationships between transcriptional control signals of the thymidine kinase gene of herpes simplex virus. *Cell* 31 (2 Pt 1): 355-65.

McMaster-Kaye, R., and J. H. Taylor. 1959. The metabolism of chromosomal ribonucleic acids in Drosophila salivary glands and its relation to synthesis of deoxyribonucleic acid. *J. Biophys. Biochem. Cytol.* 5: 461-467.

Meng, L. M., M. Kilstrup, and P. Nygaard. 1990. Autoregulation of PurR repressor synthesis and involvement of purR in the regulation of purB, purC, purL, purMN and guaBA expression in Escherichia coli. *European Journal of Biochemistry* 187 (2): 373-9.

Michalakis, Y., and M. Veuille. 1996. Length variation of CAG/CAA trinucleotide repeats in natural populations of Drosophila melanogaster and its relation to the recombination rate. *Genetics* 143 (4): 1713-25.

Min, H., C. W. Turck, J. M. Nikolic, and D. L. Black. 1997. A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes & Development* 11 (8): 1023-36.

Moens, L., J. Vanfleteren, I. De Baere, A. M. Jellie, W. Tate, and C. N. Trotman. 1992. Unexpected intron location in non-vertebrate globin genes. *FEBS Letters* 312 (2-3): 105-9.

Molina, C. A., N. S. Foulkes, E. Lalli, and P. Sassone-Corsi. 1993. Inducibility and negative autoregulation of CREM: an alternative promoter directs the expression of ICER, an early response repressor. *Cell* 75 (5): 875-86.

Moore, M. J., C. C. Query, and P. A. Sharp. 1993. Splicing of precursors to mRNAs by the spliceosome. Within *The RNA world.* R. F. Gesteland, and J. F. Atkins, pp. 303-357. Plainview, NY: Cold Spring Harbor Baboratory Press.

Mount, S. M. 1997. Genetic depletion reveals an essential role for an SR protein splicing factor in vertebrate cells. *Bioessays* 19 (3): 189-92.

Mount, S. M., C. Burks, G. Hertz, G. D. Stormo, O. White, and C. Fields. 1992. Splicing signals in Drosophila: intron size, information content, and consensus sequences. *Nucleic Acids Research* 20 (16): 4255-62.

Muller, H. J. 1935. New mutants report. *Drosophila Information Service* 3: 29-30.

Muller, M. M., T. Gerster, and W. Schaffner. 1988. Enhancer sequences and the regulation of gene transcription. *European Journal of Biochemistry* 176 (3): 485-95.

Murakami, K., M. Kimura, J. T. Owens, C. F. Meares, and A. Ishihama. 1997. The two alpha subunits of Escherichia coli RNA polymerase are asymmetrically arranged and contact different halves of the DNA upstream element. *Proceedings of the National Academy of Sciences of the United States of America* 94 (5): 1709-14.

Nagai, M., Y. Natsumeda, Y. Konno, R. Hoffman, S. Irino, and G. Weber. 1991. Selective up-regulation of type II inosine 5'-monophosphate dehydrogenase messenger RNA expression in human leukemias. *Cancer Research* 51 (15): 3886-90.

Nagai, M., Y. Natsumeda, and G. Weber. 1992. Proliferation-linked regulation of type II IMP dehydrogenase gene in human normal lymphocytes and HL-60 leukemic cells. *Cancer Research* 52 (2): 258-61.

Nagoshi, R. N., M. McKeown, K. C. Burtis, J. M. Belote, and B. S. Baker. 1988. The control of alternative splicing at genes regulating sexual differentiation in D. melanogaster. *Cell* 53 (2): 229-36.

Naguib , F. 1976. *Auxotrophic mutants of the second chromosome of Drosophila melanogaster. Ph.D. Thesis*. Edmonton, Alberta: Department of Genetics, University of Alberta.

Naguib, F. N., and D. Nash. 1976. Nucleoside auxotrophy in Drosophila: an autosomal locus yielding mutants supplementable by purine and pyrimidine ribonucleosides. *Molecular & General Genetics* 147 (1): 13-21.

Nash, D., and J. Bell. 1968. Larval age and the pattern of DNA synthesis in polytene chromosomes. *Canadian Journal of Genetics & Cytology* 10 (1): 82-90.

Nash, D., and J. F. Henderson. 1982. The biochemistry and genetics of purine metabolism in Drosophila melanogaster. *Advances in Comparative Physiology & Biochemistry* 8: 1-51.

Nash, D., and F. C. Janca. 1983. Hypomorphic lethal mutations and their implications for the interpretation of lethal complementation studies in Drosophila. *Genetics* 105: 957-968.

Nash, D., and S. Hu. 1992. Drosophila inosine monophosphate dehydrogenase is encoded at the raspberry locus. Can. Fed. Biol. Soc. Proc. 35: 72.

Nash, D., E. P. Woloshyn, Y. M. Mehl, and F. C. Janca. 1981. Pleiotropic, recessive-lethal mutants associated with purine metabolism in *Drosophila melanogaster*. *Can. J. Genet. Cytol.* 23: 411-423.

Nash, D., S. Hu, N. J. Leonard, S. Y. Tiong, and D. Fillips. 1994. The raspberry locus of Drosophila melanogaster includes an inosine monophosphate dehydrogenase like coding sequence. *Genome* 37 (2): 333-44.

Natsumeda, Y., S. Ohno, H. Kawasaki, Y. Konno, G. Weber, and K. Suzuki. 1990. Two distinct cDNAs for human IMP dehydrogenase. *Journal of Biological Chemistry* 265 (9): 5292-5.

Newlands, J. T., C. A. Josaitis, W. Ross, and R. L. Gourse. 1992. Both fis-dependent and factor-independent upstream activation of the rrnB P1 promoter are face of the helix dependent. *Nucleic Acids Research* 20 (4): 719-26.

Newman, E. B., R. D'Ari, and R. T. Lin. 1992. The leucine-Lrp regulon in E. coli: a global response in search of a raison d'etre. *Cell* 68 (4): 617-9.

Nikolov, D. B., and S. K. Burley. 1997. RNA polymerase II transcription initiation: a structural view. *Proceedings of the National Academy of Sciences of the United States of America* 94 (1): 15-22.

Nilsen, T. W. 1994. RNA-RNA interactions in the spliceosome: unraveling the ties that bind. *Cell* 78 (1): 1-4.

Niu, W., Y. Zhou, Q. Dong, Y. W. Ebright, and R. H. Ebright. 1994. Characterization of the activating region of Escherichia coli catabolite gene activator protein (CAP). I. Saturation and alanine-scanning mutagenesis. *Journal of Molecular Biology* 243 (4): 595-602.

Nolte, D. J. 1954a. The eye pigmentary system of Drosophila. IV. The pigments of the evermillion group of mutants. *J. Genet.* 52: 111-126.

Nolte, D. J. 1954b. The eye pigmentary system of Drosophila. V. The pigments of the light and dark groups of mutants. *J. Genet.* 52: 127-139.

Nolte, D. J. 1955. The eye pigmentary system of Drosophila. VI. The pigments of the ruby and red group of mutants. *J. Genet.* 53: 1-10.

Nolte, D. J. 1959. The eye pigmentary system of Drosophila. VIII. Series of multiple alleles. *Heredity* 13: 233-241.

Novina, C. D., and A. L. Roy. 1996. Core promoters and transcriptional control. *Trends in Genetics* 12 (9): 351-5.

O'Donnell, A., and D. V. Clark. Isolation of a Genomic Clone for a Purine Synthesis Gene in Drosophila. 41st Annual Meeting of the Genetics Society of Canada., London, Ontario, 1997.

O'Donnell, K. H., and P. C. Wensink. 1994. GAGA factor and TBF1 bind DNA elements that direct ubiquitous transcription of the Drosophila alpha 1-tubulin gene. *Nucleic Acids Research* 22 (22): 4712-8.

O'Donnell, K. H., C. T. Chen, and P. C. Wensink. 1994. Insulating DNA directs ubiquitous transcription of the Drosophila melanogaster alpha 1-tubulin gene. *Molecular & Cellular Biology* 14 (9): 6398-408.

Oehler, S., E. R. Eismann, H. Kramer, and B. Muller-Hill. 1990. The three operators of the lac operon cooperate in repression. *EMBO Journal* 9 (4): 973-9.

Oehler, S., M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, and B. Muller-Hill. 1994. Quality and position of the three lac operators of E. coli define efficiency of repression. *EMBO Journal* 13 (14): 3348-55.

O'Hare, K., and G. M. Rubin. 1983. Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome. *Cell* 34 (1): 25-35.

Ohshige, T., S. Takechi, and T. Nakayama. 1993. Presence of particular transcription regulatory elements in the 5'-intergenic region shared by the chicken H2A-III and H2B-V pair. *Gene* 131 (2): 193-9.

Oppenheim, D. S., and C. Yanofsky. 1980. Translational coupling during expression of the tryptophan operon of Escherichia coli. *Genetics* 95 (4): 785-95.

Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp. 1986. Splicing of messenger RNA precursors. *Annual Review of Biochemistry* 55: 1119-1150.

Palmer, M. J., V. A. Mergner, R. Richman, J. E. Manning, M. I. Kuroda, and J. C. Lucchesi. 1993. The male-specific lethal-one (msl-1) gene of Drosophila melanogaster encodes a novel protein that associates with the X chromosome in males. *Genetics* 134 (2): 545-57.

Patterson, D. 1985. *De novo* purine and pyrimidine biosynthesis. Within *Molecular cell genetics*. M. Gottesman, pp. 267-309. New York: John Wiley & Sons, Inc.

Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America* 85 (8): 2444-8.

Poole, S. J., L. M. Kauvar, B. Drees, and T. Kornberg. 1985. The engrailed locus of Drosophila: structural analysis of an embryonic transcript. *Cell* 40 (1): 37-43.

Pret, A. M., and M. Y. Fiszman. 1996. Sequence divergence associated with species-specific splicing of the nonmuscle beta-tropomyosin alternative exon. *Journal of Biological Chemistry* 271 (19): 11511-7.

Proudfoot, N. J., and G. G. Brownlee. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* 263 (5574): 211-4.

Ptashne, M. 1992. *A genetic switch: Phage [lambda] and higher organisms*. Cambridge, Mass.: Cell Press: Blackwell Scientific Publications.

Quinn, C. L., B. T. Stephenson, and R. L. Switzer. 1991. Functional organization and nucleotide sequence of the Bacillus subtilis pyrimidine biosynthetic operon. *Journal of Biological Chemistry* 266 (14): 9113-27.

Quinn, J. P., S. C. Mendelson, J. M. Paterson, J. McAllister, and C. F. Morrison. 1995. Transcriptional control of neuropeptide gene expression in sensory neurons, using the preprotachykinin-A gene as a model. *Canadian Journal of Physiology & Pharmacology* 73 (7): 957-62.

Ramseier, T. M. 1996. Cra and the control of carbon flux via metabolic pathways. *Research in Microbiology* 147 (6-7): 489-93.

Randerath, K., and E. Randerath. 1967. Thin-Layer separation methods for nucleic acid derivatives. *Methods in Enzymology* 12 (A): 323-347.

Ranish, J. A., and S. Hahn. 1996. Transcription: basal factors and activation. *Current Opinion in Genetics & Development* 6 (2): 151-8.

Rao, L., W. Ross, J. A. Appleman, T. Gaal, S. Leirmo, P. J. Schlax, M. T. Record Jr, and R. L. Gourse. 1994. Factor independent activation of rrnB P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. *Journal of Molecular Biology* 235 (5): 1421-35.

Reddy, P. S., and D. E. Housman. June 1997. The complex pathology of trinucleotide repeats. [Review] [65 refs]. *Current Opinion in Cell Biology* 9 (3): 364-72.

Reed, R. 1996. Initial splice-site recognition and pairing during pre-mRNA splicing. *Current Opinion in Genetics & Development* 6 (2): 215-20.

Reines, D., J. W. Conaway, and R. C. Conaway. 1996. The RNA polymerase II general elongation factors. *Trends in Biochemical Sciences* 21 (9): 351-5.

Rincon-Limas, D. E., R. S. Geske, J. J. Xue, C. Y. Hsu, P. A. Overbeek, and P. I. Patel. 1994. 5'-flanking sequences of the human HPRT gene direct neuronal expression in the brain of transgenic mice. *Journal of Neuroscience Research* 38 (3): 259-67.

Rincon-Limas, D. E., F. Amaya-Manzanares, M. L. Nino-Rosales, Y. Yu, T. P. Yang, and P. I. Patel. 1995. Ubiquitous and neuronal DNA-binding proteins interact with a negative regulatory element of the human hypoxanthine phosphoribosyltransferase gene. *Molecular & Cellular Biology* 15 (12): 6561-71.

Rio, D. C. 1993. Splicing of pre-mRNA: mechanism, regulation and role in development. *Current Opinion in Genetics & Development* 3 (4): 574-84.

Robberson, B. L., G. J. Cote, and S. M. Berget. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Molecular & Cellular Biology* 10 (1): 84-94.

Roeder, R. G. 1976. Eukaryotic nuclear RNA polymerases. Within *RNA polymerase*. R. Losick, and M. Chamberlin, pp. 285-329. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Roeder, R. G. 1991. The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends in Biochemical Sciences* 16 (11): 402-8.

Roeder, R. G. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences* 21 (9): 327-35.

Rogers, J. 1985. Exon shuffling and intron insertion in serine protease genes. *Nature* 315 (6019): 458-9.

Rogers, J. H. 1989. How were introns inserted into nuclear genes? *Trends in Genetics* 5 (7): 213-6.

Rokos, K., and W. Pfleiderer. 1975. New results about drosopterins. Within *Chemistry and biology of pteridines*. W. Pfleiderer, pp. 931-9. Berlin.: de Gruyter.

Rolfes, R. J., and H. Zalkin. 1988. Escherichia coli gene purR encoding a repressor protein for purine nucleotide synthesis. Cloning, nucleotide sequence, and interaction with the purF operator. *Journal of Biological Chemistry* 263 (36): 19653-61.

Rolfes, R. J., and H. Zalkin. 1990. Autoregulation of Escherichia coli purR requires two control sites downstream of the promoter. *Journal of Bacteriology* 172 (10): 5758-66.

Rolfes, R. J., and A. G. Hinnebusch. 1993. Translation of the yeast transcriptional activator GCN4 is stimulated by purine limitation: implications for activation of the protein kinase GCN2. *Molecular & Cellular Biology* 13 (8): 5099-111.

Rolfes, R. J., F. Zhang, and A. G. Hinnebusch. 1997. The transcriptional activators BAS1, BAS2, and ABF1 bind positive regulatory sites as the critical elements for adenine regulation of ADE5,7. *Journal of Biological Chemistry* 272 (20): 13343-54.

Ross, W., K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R. L. Gourse. 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262 (5138): 1407-13.

Ruppert, S. M., V. McCulloch, M. Meyer, C. Bautista, M. Falkowski, H. G. Stunnenberg, and N. Hernandez. 1996. Monoclonal antibodies directed against the amino-terminal domain of human TBP cross-react with TBP from other species. *Hybridoma* 15 (1): 55-68.

Sadowski, C. L., R. W. Henry, S. M. Lobo, and N. Hernandez. 1993. Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates transcription from snRNA promoters through the PSE. *Genes & Development* 7 (8): 1535-48.

Saier, M. H. Jr, and T. M. Ramseier. 1996. The catabolite repressor/activator (Cra) protein of enteric bacteria. *Journal of Bacteriology* 178 (12): 3411-7.

Saiki, R. K., T. L. Bugawan, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1986. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 324 (6093): 163-6.

Sakamoto, H., K. Inoue, I. Higuchi, Y. Ono, and Y. Shimura. 1992. Control of Drosophila Sex-lethal pre-mRNA splicing by its own female-specific product. *Nucleic Acids Research* 20 (21): 5533-40.

Sakashita, E., and H. Sakamoto. 1996. Protein-RNA and protein-protein interactions of the Drosophila sex-lethal mediated by its RNA-binding domains. *Journal of Biochemistry* 120 (5): 1028-33.

Salz, H. K., E. M. Maine, L. N. Keyes, M. E. Samuels, T. W. Cline, and P. Schedl. 1989. The Drosophila female-specific sex-determination gene, Sex-lethal, has stage-, tissue-, and sex-specific RNAs suggesting multiple modes of regulation. *Genes & Development* 3 (5): 708-19.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual.*, 2nd ed. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory.

Sang, J. H. 1956. The quantitative nutritional requirements of Drosophila melanogaster. *J. Exp. Biol.* 33: 45-72.

Sang, J. H. 1978. The nutritional requirements of Drosophila. Within *The genetics and biology of Drosophia.* M. Ashburner, and T. F. R. Wright, pp. 159-192. New York: Academic Press.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74 (12): 5463-7.

Sauer, R. T. 1996. Lac repressor at last. *Structure* 4 (3): 219-22.

Sawadogo, M., and A. Sentenac. 1990. RNA polymerase B (II) and general transcription factors. *Annual Review of Biochemistry* 59: 711-54.

Schafer, B., B. Wilde, D. R. Massardo, F. Manna, L. Del Giudice, and K. Wolf. 1994. A mitochondrial group-I intron in fission yeast encodes a maturase and is mobile in crosses. *Current Genetics* 25 (4): 336-41.

Schmitz, A., and D. J. Galas. 1979. The interaction of RNA polymerase and lac repressor with the lac control region. *Nucleic Acids Research* 6 (1): 111-37.

Schuler, G. D., S. F. Altschul, and D. J. Lipman. 1991. A workbench for multiple alignment construction and analysis. *Proteins* 9 (3): 180-90.

Schumacher, M. A., K. Y. Choi, H. Zalkin, and R. G. Brennan. 1994a. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science* 266 (5186): 763-70.

Schumacher, M. A., K. Y. Choi, H. Zalkin, and R. G. Brennan. 1994b. Crystallization and preliminary X-ray analysis of an Escherichia coli purine repressor-hypoxanthine-DNA complex. *Journal of Molecular Biology* 242 (3): 302-5.

Schumacher, M. A., K. Y. Choi, F. Lu, H. Zalkin, and R. G. Brennan. 1995. Mechanism of corepressor-mediated specific DNA binding by the purine repressor. *Cell* 83 (1): 147-55.

Senda, M., and Y. Natsumeda. 1994. Tissue-differential expression of two distinct genes for human IMP dehydrogenase (E.C.1.1.1.205). *Life Sciences* 54 (24): 1917-26.

Sentenac, A., M. Riva, P. Thuriaux, J. M. Buhler, I. Treich, C. Carles, M. Werner, A. Ruet, J. Huet, C. Mann, N. Chiannilkulchai, S. Stettler, and S. Mariotte. 1992. Yeast RNA polymerase subunits and genes. Within *Transcriptional regulation*. McKnight. S. L., and K. R. Yamamoto, pp. 27-54. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Serfling, E., A. Lubbe, K. Dorsch-Hasler, and W. Schaffner. 1985. Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes. *EMBO Journal* 4 (13B): 3851-9.

Shimotsu, H., M. I. Kuroda, C. Yanofsky, and D. J. Henner. 1986. Novel form of transcription attenuation regulates expression the Bacillus subtilis tryptophan operon. *Journal of Bacteriology* 166 (2): 461-71.

Shimura, K., M. Okada, H. Shiraki, and H. Nakagawa. November 1983. IMP dehydrogenase. I. Studies on regulatory properties of crude tissue extracts based on an improved assay method. *Journal of Biochemistry* 94 (5): 1595-603.

Siebenlist, U., R. B. Simpson, and W. Gilbert. 1980. E. coli RNA polymerase interacts homologously with two different promoters. *Cell* 20 (2): 269-81.

Siegfried, Z., and H. Cedar. 1997. DNA methylation: a molecular lock. *Current Biology* 7 (5): R305-7.

Sifri, C. D., K. Wilson, S. Smolik, M. Forte, and B. Ullman. 18 January 1994. Cloning and sequence analysis of a Drosophila melanogaster cDNA encoding IMP dehydrogenase. *Biochimica et Biophysica Acta* 1217 (1): 103-6.

Simpson, E. R., M. S. Mahendroo, J. E. Nichols, and S. E. Bulun. 1994. Aromatase gene expression in adipose tissue: relationship to breast cancer. *International Journal of Fertility & Menopausal Studies* 39 (Suppl 2): 75-83.

Simpson, E. R., M. D. Michael, V. R. Agarwal, M. M. Hinshelwood, S. E. Bulun, and Y. Zhao. 1997. Cytochromes P450 11: expression of the CYP19 (aromatase) gene: an unusual case of alternative promoter usage. *FASEB Journal* 11 (1): 29-36.

Slee, R., and M. Bownes. 1993. Expression of the Drosophila inosine monophosphate dehydrogenase gene. Program and abstracts, 34th Annual Drosophila Research Conference, pp. 108. San Diego, California.

Slee, R., and M. Bownes. 25 October 1995. The raspberry locus encodes Drosophila inosine monophosphate dehydrogenase. *Molecular & General Genetics* 248 (6): 755-66.

Smale, S. T., and D. Baltimore. 1989. The "initiator" as a transcription control element. *Cell* 57 (1): 103-13.

Smale, S. T., M. C. Schmidt, A. J. Berk, and D. Baltimore. 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proceedings of the National Academy of Sciences of the United States of America* 87 (12): 4509-13.

Smith, C. W., J. G. Patton, and B. Nadal-Ginard. 1989. Alternative splicing in the control of gene expression. *Annual Review of Genetics* 23: 527-77.

Smith, J. L., E. J. Zaluzec, J. P. Wery, L. Niu, R. L. Switzer, H. Zalkin, and Y. Satow. 1994. Structure of the allosteric regulatory enzyme of purine biosynthesis. *Science* 264 (5164): 1427-33.

Spofford, J. B. 1976. Position effect variegation in Drosophila. Within *The genetics and biology of Drosophila*. Vol. 1C. M. Ashburner and E. Novitski, pp. 955-1018. New York:Academic Press.

Stapel, S. O., A. Zweers, H. J. Dodemont, J. H. Kan, and W. W. de Jong. 1985. epsilon-Crystallin, a novel avian and reptilian eye lens protein. *European Journal of Biochemistry* 147 (1): 129-36.

Stehle, J. H., N. S. Foulkes, C. A. Molina, V. Simonneaux, P. Pevet, and P. Sassone-Corsi. 1993. Adrenergic signals direct rhythmic expression of transcriptional repressor CREM in the pineal gland. *Nature* 365 (6444): 314-20.

Stotz, A., P. P. Muller, and P. Linder. 1993. Regulation of the ADE2 gene from Saccharomyces cerevisiae. *Current Genetics* 24 (6): 472-80.

Sun, Z., and A. R. Means. 1995. An intron facilitates activation of the calspermin gene by the testis-specific transcription factor CREM tau. *Journal of Biological Chemistry* 270 (36): 20962-7.

Sun, Z., R. L. Means, B. LeMagueresse, and A. R. Means. 1995. Organization and analysis of the complete rat calmodulin-dependent protein kinase IV gene. *Journal of Biological Chemistry* 270 (49): 29507-14.

Szabo, G., Z. Katarova, E. Kortvely, R. J. Greenspan, and Z. Urban. 1996. Structure and the promoter region of the mouse gene encoding the 67-kD form of glutamic acid decarboxylase. *DNA & Cell Biology* 15 (12): 1081-91.

Tan, S., Y. Hunziker, D. F. Sargent, and T. J. Richmond. 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* 381 (6578): 127-51.

Tang, H., K. Severinov, A. Goldfarb, D. Fenyo, B. Chait, and R. H. Ebright. 1994. Location, structure, and function of the target of a transcriptional activator protein. *Genes & Development* 8 (24): 3058-67.

Tautz, D., and M. Renz. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* 12 (10): 4127-38.

Taylor, M. W., S. Olivelle, R. A. Levine, K. Coy, H. Hershey, K. C. Gupta, and L. Zawistowich. 1982. Regulation of de novo purine biosynthesis in Chinese hamster cells. *Journal of Biological Chemistry* 257 (1): 377-80.

Tian, M., and T. Maniatis. 1992. Positive control of pre-mRNA splicing in vitro. *Science* 256 (5054): 237-40.

Tian, M., and T. Maniatis. 1993. A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell* 74 (1): 105-14.

Tian, M., and T. Maniatis. 1994. A splicing enhancer exhibits both constitutive and regulated activities. *Genes & Development* 8 (14): 1703-12.

Tibbetts, A. S., and D. R. Appling. 1997. Saccharomyces cerevisiae expresses two genes encoding isozymes of 5-aminoimidazole-4-carboxamide ribonucleotide transformylase. *Archives of Biochemistry & Biophysics* 340 (2): 195-200.

Tiong, S. Y., and D. Nash. 1990. Genetic analysis of the adenosine3 (Gart) region of the second chromosome of Drosophila melanogaster. *Genetics* 124 (4): 889-97.

Tiong, S. Y., and D. Nash. 1993. The adenosine2 gene of Drosophila melanogaster encodes a formylglycineamide ribotide amidotransferase. *Genome* 36 (5): 924-34.

Tiong, S. Y., C. Keizer, D. Nash, J. Bleskan, and D. Patterson. June 1989. Drosophila purine auxotrophy: new alleles of adenosine 2 exhibiting a complex visible phenotype. *Biochemical Genetics* 27 (5-6): 333-48.

Tjian, R., and T. Maniatis. 1994. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 77 (1): 5-8.

Toda, K., S. Nomoto, and Y. Shizuta. 1996. Identification and characterization of transcriptional regulatory elements of the human aromatase cytochrome P450 gene (CYP19). *Journal of Steroid Biochemistry & Molecular Biology* 56 (1-6 Spec No): 151-9.

Torres, M., and L. Sanchez. 1992. The segmentation gene runt is needed to activate Sex-lethal, a gene that controls sex determination and dosage compensation in Drosophila. *Genetical Research* 59 (3): 189-98.

Tsuji, S. 1997. Molecular genetics of triplet repeats: unstable expansion of triplet repeats as a new mechanism for neurodegenerative diseases. *Internal Medicine* 36 (1): 3-8.

Ushida, C., and H. Aiba. 1990. Helical phase dependent action of CRP: effect of the distance between the CRP site and the -35 region on promoter activity. *Nucleic Acids Research* 18 (21): 6325-30.

Valcarcel, J., and M. R. Green. 1996. The SR protein family: pleiotropic functions in pre-mRNA splicing. *Trends in Biochemical Sciences* 21 (8): 296-301.

Verham, R., T. D. Meek, L. Hedstrom, and C. C. Wang. May 1987. Purification, characterization, and kinetic analysis of inosine 5'-monophosphate dehydrogenase of Tritrichomonas foetus. *Molecular & Biochemical Parasitology* 24 (1): 1-12.

Verrijzer, C. P., and R. Tjian. 1996. TAFs mediate transcriptional activation and promoter selectivity. *Trends in Biochemical Sciences* 21 (9): 338-42.

Verrijzer, C. P., J. L. Chen, K. Yokomori, and R. Tjian. 1995. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* 81 (7): 1115-25.

Vieira, J., and J. Messing. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19 (3): 259-68.

Villee, C. A., and H. B. Bissell. 1948. Nucleic acids as growth factors in Drosophila. *J. Biol. Chem.* 172: 59-66.

Voet, D., and J. G. Voet. 1995. *Biochemistry.* New York: John Wiley & Sons, Inc.

Vossen, K. M., D. F. Stickle, and M. G. Fried. 1996. The mechanism of CAP-lac repressor binding cooperativity at the E. coli lactose promoter. *Journal of Molecular Biology* 255 (1): 44-54.

Vyse, E. R., and D. Nash. 1969. Nutritional conditional mutants of Drosophila melanogaster. *Genetical Research* 13 (3): 281-7.

Vyse, E. R., and J. H. Sang. 1971. A purine and pyrimidine requiring mutant of Drosophila melanogaster. *Genetical Research* 18 (1): 117-21.

Watakabe, A., K. Tanaka, and Y. Shimura. 1993. The role of exon sequences in splice site selection. *Genes & Development* 7 (3): 407-18.

Weber, G. 1983. Biochemical strategy of cancer cells and the design of chemotherapy: G.H.A. Clowes Memorial Lecture. *Cancer Res* 43:3466-3492.

Weber, I. T., and T. A. Steitz. 1987. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 A resolution. *Journal of Molecular Biology* 198 (2): 311-26.

Weil, P. A., D. S. Luse, J. Segall, and R. G. Roeder. 1979. Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. *Cell* 18 (2): 469-84.

Weis, L., and D. Reinberg. 1992. Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB Journal* 6 (14): 3300-9.

Wells, J., P. Held, S. Illenye, and N. H. Heintz. 1996. Protein-DNA interactions at the major and minor promoters of the divergently transcribed dhfr and rep3 genes during the Chinese hamster ovary cell cycle. *Molecular & Cellular Biology* 16 (2): 634-47.

Weng, M., P. L. Nagy, and H. Zalkin. 1995. Identification of the Bacillus subtilis pur operon repressor. *Proceedings of the National Academy of Sciences of the United States of America* 92 (16): 7455-9.

Werner, E. R., G. Werner-Felmayer, H. Wachter, and B. Mayer. 1996. Biosynthesis of nitric oxide: dependence on pteridine metabolism. *Reviews of Physiology Biochemistry & Pharmacology* 127: 97-135.

Wharton, K. A., B. Yedvobnick, V. G. Finnerty, and S. Artavanis-Tsakonas. January 1985. opa: a novel family of transcribed repeats shared by the Notch locus and other developmentally regulated loci in D. melanogaster. *Cell* 40 (1): 55-62.

White, J. A., W. Liu, and J. E. Wilson. 1996. Isolation of the promoter for Type I hexokinase from rat. *Archives of Biochemistry & Biophysics* 335 (1): 161-72.

Wieschaus, E., and C. Nusslein-Vollhard. 1986. Looking at embryos. Within *Drosophila, A practical approach*. D. B. Roberts, pp. 199-228. Oxford: IRL Press.

Will, C. L., and R. Luhrmann. 1997. Protein functions in pre-mRNA splicing. *Current Opinion in Cell Biology* 9 (3): 320-8.

Williams, J. A., S. W. Paddock, K. Vorwerk, and S. B. Carroll. 1994. Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. *Nature* 368 (6469): 299-305.

Wistow, G. J., J. W. Mulders, and W. W. de Jong. 1987. The enzyme lactate dehydrogenase as a structural protein in avian and crocodilian lenses. *Nature* 326 (6113): 622-4.

Wu, J. Y., and T. Maniatis. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75 (6): 1061-70.

Wyss, C. 1977. Purine and pyrimidine salvage in a clonal Drosophila cell line. *Journal of Insect Physiology* 23 (6): 739-47.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33 (1): 103-19.

Yanofsky, C. 1992. Transcriptional regulation: Elegance in design and discovery. Within *Transcriptional regulation*. S. L. McKnight, and K. R. Yamamoto, pp. 3-24. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Yim, J. J., E. H. Grell, and K. B. Jacobson. 1977. Mechanism of suppression in Drosophila: control of sepiapterin synthase at the purple locus. *Science* 198 (4322): 1168-70.

Young, R. A. 1991. RNA polymerase II. *Annual Review of Biochemistry* 60: 689-715.

Younger-Shepherd, S., H. Vaessin, E. Bier, L. Y. Jan, and Y. N. Jan. 1992. deadpan, an essential pan-neural gene encoding an HLH protein, acts as a denominator in Drosophila sex determination. *Cell* 70 (6): 911-22.

Zahler, A. M., W. S. Lane, J. A. Stolk, and M. B. Roth. 1992. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes & Development* 6 (5): 837-47.

Zalkin, H. 1993. Overview of multienzyme systems in biosynthetic pathways. *Biochemical Society Transactions* 21 (1): 203-7.

Zalkin, H., and D. J. Ebbole. 1988. Organization and regulation of genes encoding biosynthetic enzymes in Bacillus subtilis. *Journal of Biological Chemistry* 263 (4): 1595-8.

Zalkin, H., and J. E. Dixon. 1992. *De Novo* Purine Nucleotide Biosynthesis. *Progress in Nucleic Acid Research and Molecular Biology* 42: 259-287.

Zawel, L., and D. Reinberg. 1993. Initiation of transcription by RNA polymerase II: a multi-step process. *Progress in Nucleic Acid Research & Molecular Biology* 44: 67-108.

Zawel, L., and D. Reinberg. 1995. Common themes in assembly and function of eukaryotic transcription complexes. *Annual Review of Biochemistry* 64: 533-61.

Zhang, F., M. Kirouac, N. Zhu, A. G. Hinnebusch, and R. J. Rolfes. 1997. Evidence that complex formation by Bas1p and Bas2p (Pho2p) unmasks the activation function of Bas1p in an adenine-repressible step of ADE gene transcription. *Molecular & Cellular Biology* 17 (6): 3272-83.

Zhou, G., J. L. Smith, and H. Zalkin. 1994. Binding of purine nucleotides to two regulatory sites results in synergistic feedback inhibition of glutamine 5-phosphoribosylpyrophosphate amidotransferase. *Journal of Biological Chemistry* 269 (9): 6784-9.

Zhou, S., Y. Yang, M. J. Scott, A. Pannuti, K. C. Fehr, A. Eisen, E. V. Koonin, D. L. Fouts, R. Wrightsman, J. E. Manning et al. 1995. Male-specific lethal 2, a dosage compensation gene of Drosophila, undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster. *EMBO Journal* 14 (12): 2884-95.

Zhou, Y., S. Busby, and R. H. Ebright. 1993. Identification of the functional subunit of a dimeric transcription activator protein by use of oriented heterodimers. *Cell* 73 (2): 375-9.

Zhou, Y., T. J. Merkel, and R. H. Ebright. 1994. Characterization of the activating region of Escherichia coli catabolite gene activator protein (CAP). II. Role at Class I and class II CAP-dependent promoters. *Journal of Molecular Biology* 243 (4): 603-10.

Zhu, C., J. Urano, and L. R. Bell. 1997. The Sex-lethal early splicing pattern uses a default mechanism dependent on the alternative 5' splice sites. *Molecular & Cellular Biology* 17 (3): 1674-81.

Zhu, Q. S., K. Chen, and J. C. Shih. 1994. Bidirectional promoter of human monoamine oxidase A (MAO A) controlled by transcription factor Sp1. *Journal of Neuroscience* 14 (12): 7393-403.

Ziegler, I. 1961. Genetic aspects of ommochrome and pterin pigments. *Adv. in Genet.* 10: 349-403.

Ziegler, I., and R. Harmsen. 1969. The biology of pteridines in insects. *Adv. Insect Physiol.* 6: 139-203.

Zimmermann, A. G., J. Spychala, and B. S. Mitchell. 1995. Characterization of the human inosine-5'-monophosphate dehydrogenase type II gene. *Journal of Biological Chemistry* 270 (12): 6808-14.

Zimmermann, A., J. J. Gu, J. Spychala, and B. S. Mitchell. 1996. Inosine monophosphate dehydrogenase expression: transcriptional regulation of the type I and type II genes. *Advances in Enzyme Regulation* 36: 75-84.

Zubay, G., D. Schwartz, and J. Beckwith. 1970. Mechanism of activation of catabolite-sensitive genes: a positive control system. *Proceedings of the National Academy of Sciences of the United States of America* 66 (1): 104-10.

Appendices

Appendix 1 Protein concentration of crude extracts

A. The standard curve of protein concentration

	1st sample	2nd sample	Average reading	Converted concentration
1 ug/ml	0.077	0.067	0.072	1.010
2 ug/ml	0.142	0.156	0.149	2.022
3 ug/ml	0.231	0.200	0.216	3.008
4 ug/ml	0.289	0.257	0.273	3.945
5 ug/ml	0.349	0.313	0.331	4.968
6 ug/ml	0.377	0.380	0.379	5.866
7 ug/ml	0.438	0.440	0.439	7.086
8 ug/ml	0.496	0.442	0.469	7.723
9 ug/ml	0.532	0.528	0.530	9.084
10 ug/ml	0.566	0.555	0.561	9.797

The protein standards were read by UV spectrophotometer as described in the text. The standard curve was generated using Cricket Graph III software.

B. Experimental samples

	Samples	1 ul	1 ul	Ave 1 ul	Protein concentration (ug/ul)
Am^+ female	La	0.739	0.703	0.721	13.911
	P1	0.49	0.467	0.479	7.929
	P2	0.536	0.506	0.521	8.877
	P3	0.414	0.386	0.400	6.290
	P3.5	0.534	0.483	0.509	8.594
	P4	0.405	0.375	0.390	6.091
	YA	0.538	0.533	0.536	9.211
Am^+ male	La	0.595	0.65	0.623	11.314
	P1	0.379	0.425	0.402	6.330
	P2	0.403	0.328	0.366	5.615
	P3	0.375	0.342	0.359	5.482
	P3.5	0.432	0.451	0.442	7.138
	P4	0.392	0.339	0.366	5.615
	YA	0.472	0.457	0.465	7.626
ras^1 female	La	0.537	0.468	0.503	8.459
	P1	0.745	0.674	0.710	13.596
	P2	0.624	0.648	0.636	11.656
	P3	0.573	0.581	0.577	10.192
	P4	0.576	0.474	0.525	8.969
	YA	0.748	0.586	0.667	12.459

ras^2 female	La	0.589	0.548	0.569	9.987
	P1	0.776	0.641	0.709	13.569
	P2	0.56	0.483	0.522	8.889
	P3	0.464	0.439	0.452	7.349
	P4	0.458	0.435	0.447	7.243
	YA	0.558	0.537	0.548	9.490
ras^3 female	La	0.661	0.584	0.623	11.314
	P1	0.537	0.541	0.539	9.292
	P2	0.507	0.479	0.493	8.248
	P3	0.44	0.498	0.469	7.723
	P3.5	0.46	0.449	0.455	7.412
	P4	0.569	0.562	0.566	9.916
	YA	0.512	0.482	0.497	8.337
ras^3 male	La	0.55	0.49	0.520	8.855
	P1	0.502	0.46	0.481	7.984
	P2	0.39	0.364	0.377	5.837
	P3	0.391	0.371	0.381	5.915
	P3.5	0.38	0.349	0.365	5.596
	P4	0.452	0.385	0.419	6.663
	YA	0.376	0.368	0.372	5.740
ras^4 female	YA	0.445	0.418	0.432	6.930

Appendix 2 IMPDH activity in late larval to young adult stages of wild-type (Am^+) strain

Samples	Density	Area	Activity	Protein ug/ml	Activity/ug
La-3 min.	46954	15855	17168	8.871	1935
La-6 min.	51178	15855	21392	8.871	2411
La-9 min.	52642	15704	23139	8.871	2608
P1-3 min.	75235	15855	45448	7.887	5762
P1-6 min.	108150	15855	78363	7.887	9935
P1-9 min.	118708	15855	88921	7.887	11274
P2-3 min.	57555	14994	29386	8.399	3499
P2-6 min.	80972	14847	53079	8.399	6320
P2-9 min.	94485	14994	66316	8.399	7896
P3-3 min.	43521	14994	15352	6.246	2458
P3-6 min.	54364	14847	26471	6.246	4238
P3-9 min.	61222	14994	33053	6.246	5292
P3.5-3 min.	59650	17762	26280	7.999	3285
P3.5-6 min.	75211	17596	42154	7.999	5270
P3.5-9 min.	81756	17762	48387	7.999	6049
P4-3 min.	63567	17762	30197	6.051	4990
P4-6 min.	89737	17596	56680	6.051	9367
P4-9 min.	107129	17762	73760	6.051	12189
YA-3 min.	68780	14994	40611	8.388	4842
YA-6 min.	91740	15141	63294	8.388	7546
YA-9 min.	93364	14994	65195	8.388	7773
Background	20365	10840	0		

The table is generated from the analysis of the PhosphorImager scanned data of the ^{14}C-XMP signal on the TLC plate in Figure 3-22 by ImageQuant software. The calculated areas were manually selected depending on the size of the signals. The background was selected on an area without signals. In the table,

D^I (Density of IMPDH signal): sum of all pixel values in the selected area.
A^I (Area of IMPDH signal): total number of pixels in the selected area.
D^B (Density of Background): sum of all pixel values in background area.
A^B (Area of Background): total number of pixels in background area.
Activity $D^I - A^I \times (D^B \div A^B)$
Ativity/µg: Activity ÷ protein concentration.

Appendix 3 IMPDH activity in the late developmental stage

Samples: LA	Density	Area	Activity	Protein ug/ml	Activity/ug
WT female-3 min.	87479	15600	45939	8.871	5178
WT female-6 min.	100137	15450	58996	8.871	6650
WT female-9 min.	106399	15600	64859	8.871	7311
WT male-3 min.	108986	15450	67845	7.754	8750
WT male-6 min.	142428	15600	100887	7.754	13011
WT male-9 min.	177441	15450	136300	7.754	17578
ras^1 female-3 min.	114752	18054	66677	7.865	8478
ras^1 female-6 min.	131894	17877	84290	7.865	10717
ras^1 female-9 min.	138699	18054	90624	7.865	11522
ras^2 female-3 min.	100673	17877	53069	9.422	5632
ras^2 female-6 min.	117104	18054	69029	9.422	7326
ras^2 female-9 min.	127040	17877	79436	9.422	8431
ras^3 female-3 min.	147031	20988	91144	9.649	9446
ras^3 female-6 min.	174032	20988	118144	9.649	12245
ras^3 female-9 min.	179733	20790	124373	9.649	12890
ras^3 male-3 min.	139621	20988	83734	8.847	9465
ras^3 male-6 min.	181902	20988	126014	8.847	14244
ras^3 male-9 min.	200949	20988	145061	8.847	16396
background	30862	11590	0		

Samples: P1	Density	Area	Activity	Protein ug/ml	Activity/ug
WT female-3 min.	83036	18480	32456	7.887	4115
WT female-6 min.	93868	18480	43288	7.887	5488
WT female-9 min.	96796	18656	45735	7.887	5798
WT male-3 min.	74231	18480	23650	6.285	3763
WT male-6 min.	85485	18480	34905	6.285	5554
WT male-9 min.	92487	18480	41907	6.285	6668
ras^1 female-3 min.	56360	19588	2747	9.199	299
ras^1 female-6 min.	59037	19588	5425	9.199	590
ras^1 female-9 min.	59955	19422	6797	9.199	739
ras^2 female-3 min.	80670	19588	27057	9.802	2761
ras^2 female-6 min.	107165	19588	53552	9.802	5464
ras^2 female-9 min.	116267	19588	62654	9.802	6392
ras^3 female-3 min.	55586	19602	1935	9.310	208
ras^3 female-6 min.	56673	19602	3022	9.310	325
ras^3 female-9 min.	57298	19440	4090	9.310	439
ras^3 male-3 min.	70720	19602	17068	7.943	2149
ras^3 male-6 min.	91270	19602	37619	7.943	4736
ras^3 male-9 min.	104139	19602	50488	7.943	6356
background	15264	5577	0		

The tables in Appendix 3 are generated from the data in Figure 3-24 (LA, P1, P2, P3, P4 and LA) by using the method described in Appendix 2 respectively.

Samples: P2	Density	Area	Activity	Protein ug/ml	Activity/ug
WT female-3 min.	196403	22826	140086	8.399	16679
WT female-6 min.	270660	23028	213844	8.399	25460
WT female-9 min.	345706	22826	289388	8.399	34454
WT male-3 min.	172158	22826	115841	5.587	20733
WT male-6 min.	233557	23028	176741	5.587	31633
WT male-9 min.	295271	22826	238954	5.587	42768
ras^{1} female-3 min.	236017	24955	174446	8.056	21655
ras^{1} female-6 min.	305941	24738	244905	8.056	30401
ras^{1} female-9 min.	360786	24955	299216	8.056	37143
ras^{2} female-3 min.	210655	24738	149620	8.883	16843
ras^{2} female-6 min.	250747	24955	189176	8.883	21296
ras^{2} female-9 min.	285117	24738	224082	8.883	25226
ras^{3} female-3 min.	92555	22422	37234	8.215	4533
ras^{3} female-6 min.	98711	22624	42892	8.215	5221
ras^{3} female-9 min.	101013	22422	45693	8.215	5562
ras^{3} male-3 min.	78080	22422	22759	5.803	3922
ras^{3} male-6 min.	89500	22624	33681	5.803	5804
ras^{3} male-9 min.	92226	22422	36905	5.803	6360
background	54823	22220	0		

Samples:P3	Density	Area	Activity	Protein ug/ml	Activity/ug
WT female-3 min.	70988	16352	30577	6.246	4896
WT female-6 min.	87640	16352	47229	6.246	7562
WT female-9 min.	105448	16206	65398	6.246	10471
WT male-3 min.	59208	16352	18797	5.458	3444
WT male-6 min.	67539	16352	27127	5.458	4970
WT male-9 min.	77962	16352	37551	5.458	6880
ras^{1} female-3 min.	97103	19488	48941	8.422	5811
ras^{1} female-6 min.	127979	19488	79818	8.422	9477
ras^{1} female-9 min.	143240	19656	94664	8.422	11240
ras^{2} female-3 min.	77206	19488	29045	7.299	3980
ras^{2} female-6 min.	88009	19488	39848	7.299	5460
ras^{2} female-9 min.	91784	19488	43622	7.299	5977
ras^{3} female-3 min.	74371	20826	22903	7.677	2983
ras^{3} female-6 min.	83129	20648	32101	7.677	4181
ras^{3} female-9 min.	98241	20826	46773	7.677	6093
ras^{3} male-3 min.	62128	20826	10661	5.879	1813
ras^{3} male-6 min.	62983	20648	11955	5.879	2034
ras^{3} male-9 min.	64206	20826	12738	5.879	2167
background	28489	11528	0		

Samples: P4	Density	Area	Activity	Protein ug/ml	Activity/ug
WT female-3 min.	114788	19184	68654	6.051	11345
WT female-6 min.	138433	19184	92299	6.051	15253
WT female-9 min.	159230	19008	113519	6.051	18759
WT male-3 min.	101871	19184	55737	5.587	9976
WT male-6 min.	122100	19184	75965	5.587	13596
WT male-9 min.	140617	19184	94482	5.587	16910
ras^1 female-3 min.	80843	18792	35651	8.376	4256
ras^1 female-6 min.	99891	18792	54700	8.376	6531
ras^1 female-9 min.	111232	18954	65651	8.376	7838
ras^2 female-3 min.	56352	18792	11160	7.192	1552
ras^2 female-6 min.	59431	18792	14239	7.192	1980
ras^2 female-9 min.	63129	18792	17938	7.192	2494
ras^3 female-3 min.	84668	21060	34022	9.224	3689
ras^3 female-6 min.	103480	21240	52401	9.224	5681
ras^3 female-9 min.	114029	21060	63383	9.224	6872
ras^3 male-3 min.	60890	21060	10245	6.614	1549
ras^3 male-6 min.	65503	21240	14424	6.614	2181
ras^3 male-9 min.	67216	21060	16570	6.614	2505
background	41618	17306	0		

Samples: YA	Density	Area	Activity	Protein ug/ml	Activity/ug
WT female-3 min.	66003	14949	32934	8.388	3927
WT female-6 min.	77479	14949	44410	8.388	5295
WT female-9 min.	82949	14798	50214	8.388	5987
WT male-3 min.	54662	14949	21593	7.579	2849
WT male-6 min.	65604	14949	32536	7.579	4293
WT male-9 min.	68208	14949	35140	7.579	4637
ras^1 female-3 min.	37740	16006	2333	9.016	259
ras^1 female-6 min.	41056	15855	5984	9.016	664
ras^1 female-9 min.	94273	16006	58866	9.016	6529
ras^2 female-3 min.	38002	16006	2595	9.522	272
ras^2 female-6 min.	38718	15855	3646	9.522	383
ras^2 female-9 min.	40626	16006	5219	9.522	548
ras^3 female-3 min.	39286	16157	3545	8.307	427
ras^3 female-6 min.	39982	16157	4241	8.307	511
ras^3 female-9 min.	40319	16157	4578	8.307	551
ras^3 male-3 min.	39010	16157	3269	6.685	489
ras^3 male-6 min.	41323	16157	5582	6.685	835
ras^3 male-9 min.	45270	16157	9529	6.685	1426
ras^4 female-3 min.	66279	13299	36860	6.275	5874
ras^4 female-6 min.	88157	13442	58421	6.275	9310
ras^4 female-9 min.	92407	13299	62989	6.275	10038
background	14958	6762	0		

Appendix 4 Wild-type IMPDH RNA levels in 0-88 hour Drosophila embryos

	MPDH			RP49			IMPDH/RP49
Samples	Area	Density	Adjusted D	Area	Density	Adjusted D	Relative D (%)
0-2 hour	380	289	190	403	143	37	520
2-4 hour	36	13	4	423	160	47	9
4-8 hour	67	22	4	363	149	52	8
8-16 hour	144	98	61	461	207	85	72
16-20 hour	97	65	40	565	237	87	46
20-40 hour	65	36	19	565	237	88	21
40-64 hour	92	62	38	652	249	76	50
64-88 hour	67	38	20	636	208	40	52
background	170	44	0	440	117	0	0

The table is generated from the northern data shown in Figure 3-26. The calculated areas were manually selected depending on the intensity of the signals. The background was selected on an area without signals. In the table:

D^I (Density of IMPDH signal): sum of all pixel values in the selected area.
A^I (Area of IMPDH signal): total number of pixels in the selected area.
D^B (Density of Background): sum of all pixel values in background area.
A^B (Area of Background): total number of pixels in background area.
D^R (Density of RP49 signal): sum of all pixel values in the selected area.
A^R (Area of RP49 signal): total number of pixels in the selected area.
D^{AI} (Adjusted D^I): $D^I - A^I \times (D^B \div A^B)$
D^{AR} (Adjusted D^R): $D^R - A^R \times (D^B \div A^B)$
Relative D: $D^{AI} \div D^{AR}$ X 100

Appendix 5 Wild-type IMPDH RNA levels from late larval to young adult stage

	IMPDH			RP49			IMPDH/RP49
Samples	Area	Density	relative D	Area	Density	Relative D	Adjusted D (%)
La Am$^+$ female	164	126	85	157	128	85	100
La OrR female	136	108	74	142	125	87	86
La Am$^+$ male	186	161	115	119	91	58	197
La OrR male	190	166	119	120	106	74	161
P1 female	149	114	77	105	79	50	154
P1 male	143	107	71	90	51	27	264
P2 female	107	70	44	98	62	35	125
P2 male	110	75	48	100	56	29	166
P3 female	116	92	63	89	53	29	216
P3 male	97	66	42	106	73	44	94
P3.5 female	93	58	35	111	92	62	56
P3.5 male	68	31	14	111	87	57	25
P4 female	113	97	69	160	127	84	82
P4 male	94	71	47	135	111	75	63
YA female	146	132	96	157	129	87	111
YA male	217	187	133	157	128	86	155
background	371	92	0	177	48	0	0

The table is generated from the northern data shown in Figure 3-27 by using the method described in Appendix 4.

Appendix 6 Northern analysis of *ras*^{1} late larvae to young adult samples

	IMPDH			RP49			IMPDH/RP49
Samples	Area	Density	relative D	Area	Density	Relative D	Adjusted D (%)
La female	87	36	14	148	118	80	18
La male	81	45	24	163	134	91	27
P1 female	113	78	49	121	102	70	70
P1 male	129	96	63	116	98	67	94
P2 female	118	97	67	120	106	74	90
P2 male	118	81	51	135	113	78	65
P3 female	83	27	6	102	70	43	14
P3 male	84	39	17	122	93	61	29
P4 female	103	65	38	137	114	78	49
P4 male	80	41	20	99	91	66	31
YA female	133	102	68	102	81	54	125
YA male	161	135	93	122	109	77	122
background	182	47	0	267	70	0	0

The table is generated from the northern data shown in Figure 3-30 by using the method described in Appendix 4.

Appendix 7 Northern analysis of *ras*^{2} late larvae to young adult samples

	IMPDH			RP49			IMPDH/RP49
Samples	Area	Density	relative D	Area	Density	Relative D	Adjusted D (%)
La female	100	42	19	190	133	82	23
La male	72	38	22	182	136	87	25
P1 female	107	75	50	179	138	90	56
P1 male	68	38	22	166	103	59	38
P2 female	78	39	21	153	102	61	35
P2 male	52	17	6	169	113	68	8
P3 female	109	50	26	136	84	48	53
P3 male	88	42	22	145	112	74	30
P4 female	88	37	17	152	114	74	23
P4 male	67	23	8	133	77	42	19
YA female	200	161	116	234	178	115	100
YA male	154	122	87	211	149	93	93
background	395	90	0	294	78	0	0

The table is generated from the northern data shown in Figure 3-32 by using the method described in Appendix 4.

Appendix 8 Northern analysis of *ras*^{3} late larvae to young adult and *ras*^{4} young adult samples

	IMPDH			RP49			IMPDH/RP49
Samples *ras*^{3}	Area	Density	relative D	Area	Density	Relative D	Adjusted D (%)
La female	173	149	105	173	142	101	104
La male	122	76	45	161	147	109	42
P1 female	94	40	17	159	128	91	19
P1 male	101	47	22	127	97	67	32
P2 female	102	33	7	91	59	38	19
P2 male	82	38	17	100	60	37	47
P3 female	115	47	19	91	58	36	51
P3 male	114	37	8	97	67	44	19
P4 female	113	57	29	101	74	50	57
P4 male	83	39	18	87	71	51	36
YA female	167	143	101	138	112	79	128
YA male	175	142	98	120	96	68	146
Samples *ras*^{4}							
YA female	164	157	116	151	128	92	126
YA male	151	129	91	184	135	91	100
background	219	55	0	323	77	0	0

The table is generated from the northern data shown in Figure 3-34 by using the method described in Appendix 4.

Appendix 9 Northern analysis of 0-88 hour Am$^+$ and *ras*^{2} samples

	IMPDH			RP49			IMPDH/RP49
Samples	Area	Density	adjusted D	Area	Density	Adjusted D	Relative D (%)
OrR 0-2 hours	380	288	190	403	143	37	520
2-4 hour	36	13	4	423	159	47	9
4-8 hour	67	22	4	363	149	52	8
8-16 hour	144	98	61	461	207	85	72
16-20 hour	97	65	40	565	237	87	46
20-40 hour	65	35	19	565	237	88	21
40-64 hour	92	62	38	652	249	76	50
64-88 hour	67	38	20	636	208	40	51
ras^2 0-2 hour	82	27	6	322	113	27	22
2-4 hour	87	28	5	359	122	26	21
4-8 hour	71	23	5	242	71	7	71
8-16 hour	146	111	73	439	165	49	149
16-20 hour	51	29	16	432	161	46	34
20-40 hour	190	175	126	680	269	88	143
40-64 hour	66	22	4	545	193	48	9
64-88 hour	76	21	1	579	187	33	4
background	170	44	0	440	117	0	0

The table is generated from the northern data shown in Figure 3-36 by using the method described in Appendix 4.